INDIAN STATISTICAL INSTITUTE

Periodical Examination

M. Tech (CS) - II

Computational geometry

Date : 10.09.2018 Maximum Marks : 30

Duration : 2 hours.

(Although the paper carries a total marks of 40, the maximum marks you can score is 30)

Question 1:(a) State and prove the lower-bound on the running time of Intersection of n line segments. 5

(b) Describe Chan's algorithm for computing Convex hull of P. 5

Question 2: (a) If the center of minimum enclosing disk is constrained to lie on the x-axis then which are the points, in Figure 1, will get pruned after first iteration for the given pairing of the points. The bisectors of the pairs are shown in the figure. 5

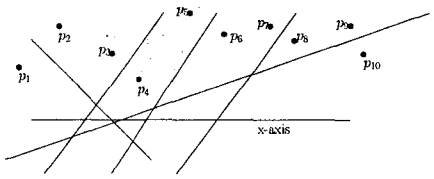


Figure 1:

(b) List the set of points that will be get deleted from STACK while running Graham Scan algorithm for computing the upper envelope of the convex hull of the point set shown in Figure 2. 5

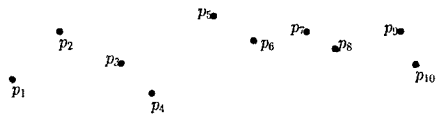


Figure 2:

Question 3: (a) Let S be a convex polygon of n vertices. Construct data structures, that use $O(n)$ storage and time, that can answer the following query in $O(\log n)$ time: given any query line l whether l intersects S. 5

(b) Show that, the dual graph of a triangulated simple polygon is a tree. 5

Question 4: (a) Let P and Q are two convex polygon with n and m vertices, respectively. Write a $O(n+m)$ time algorithm to find $P \bigcap Q$. 5

(b) Use above algorithm to find intersection of n half planes. 5

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination : (2018 - 2019)

Course Name : M. Tech. (CS) **Year : 2nd year**

Subject Name : Neural Networks and Applications

Date : September 10, 2018 **Maximum Marks : 50** **Duration : 2 hrs**

Answer all the questions.

1. Define an artificial neural network. What are its advantages over a traditional computing system? [2 + 5 = 7]

2. State and prove perceptron convergence theorem. [5 + 15 = 20]

3. Consider a single layered artificial neural network with a single neuron following McCulloch-Pitts model. The activation function of the neuron is $f(x) = x$, where x is the total input received by the neuron. The neuron accepts n-dimensional input vectors, but no bias, through an n-dimensional weight vector. The network learns under Hebb rule of learning such that the weight vector is normalized after each iteration.

   (i) Show how the weight vector, after convergence, can land up to the first principal component of the data comprising these n-dimensional input vectors.

   (ii) Show how the network be augmented to extract the fourth principal component of the data. [20 + 5 = 25]

INDIAN STATISTICAL INSTITUTE

Periodical Examination

M. Tech (CS) - I Year (Semester - I)

Optimization Technique

Date : 11.09.2018 Maximum Marks : 60 Duration : 3 Hours

Note : You may answer any part of any question, but maximum you can score is 60.

1. Consider a set of points $P = \{p_1, p_2, \ldots, p_n\}$ in R^2. In other words, the coordinate of the points in P are the input of the problem. The objective is to place non-overlapping circles (may be of different radii) centered at those points such that the sum of perimeter of the placed circles is maximized. Formulate the problem as a linear programming problem. [10]

2. Define basic frasible solution (BFS) of an LP problem.
   (a) Show that, if for a set of m linear equations with n unknowns ($m \leq n$), there is a feasible solution, then there always exists a basic feasible solution.
   (b) Show that every BFS is an extreme point of the convex polyhedron defined by the constraints of the LP problem. [3+9+8=20]

3. Solve the following linear programming problem

$$\max z = 2x_1 + 3x_2 + x_3$$

subject to

$$4x_1 + 3x_2 + x_3 \leq 6$$
$$x_1 + 2x_2 + 5x_3 \leq 4$$
$$x_1, x_2, x_3 \geq 0$$

[10]

4. Consider the following 1-variable linear programming problem, which we call P:

$$\begin{aligned} \text{maximize} \quad & z = tx \\ \text{subject to:} \quad & rx \leq s \\ & x \geq 0 \end{aligned}$$

where r, s and t are arbitrary real numbers. Let D be the dual of P.

State for which values of r, s and t you can assert that

A. Both P and D have optimal solutions with finite objective values.

B. P is feasible, but D is not feasible.

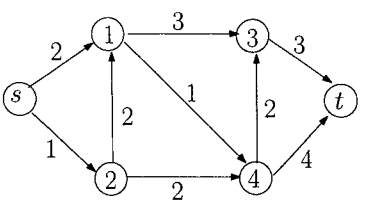
C. D is feasible, but P is not feasible.

D. Neither P not D is feasible.

. $[2.5 \times 4 = 10]$

5. (a) Write down the *complementary slackness conditions.*

(b) Use it to develope the primal dual algorithm for computing the shortest path problem between the given pair of nodes s and t of the following edge-weighted directed graph. Write down the steps of your algorithm.



(c) Design a primal dual algorithm for designing a f factor approximation algorithm for the set cover problem with the universe $U = \{e_1, e_2, \ldots, e_n\}$ of elements and given subsets $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$, where $S_i \subseteq U$ and f is the maximum number of occurance of an element $e \in U$ in the subsets of S. [5+10+10=25]

Indian Statistical Institute
Mid-Semestral Examination: 2018
Course Name: M. Tech. in Computer Science
Subject Name: Mobile Computing

Date: 11-09-2018 Maximum Marks: 60 Duration: 3 hours

Instructions: You **may** attempt **all** questions which carry a total of **65** marks. However, the maximum marks you can score is only **60**.

1. (a) What is the difference between *soft* handover and *hard* handover? [1]
   (b) What is the difference between *horizontal* handover and *vertical* handover? [1]
   (c) What is *network connection time*? [1]
   (d) What are *unnecessary* handover and *superfluous* handover? [2]
   (e) State two main reasons for handover failure. [2]
   (f) State the three stages of a handover process. [2]
   (g) Briefly describe the *adaptive lifetime* based vertical handover decision algorithm. [5]
   (h) Briefly describe a *cost function* based vertical handover decision algorithm. [5]
   (i) Explain the *relative signal strength with threshold* based hard handover strategy. [3]

2. (a) What are *reporting cells* and *location areas* in static location update schemes? [4]
   (b) Describe the *time*, *movement* and *distance* based dynamic location update schemes. [6]
   (c) State the main difference between *simultaneous* paging and *sequential* paging. [3]
   (d) Briefly describe the *random waypoint* mobility model for characterizing individual node movement. [4]
   (e) Briefly describe the *community* based mobility model for characterizing the movement of groups of nodes. [4]
   (f) State the *skewed location preference* and *periodical reappearance* properties of a mobile node. [2]

3. (a) Represent the *channel assignment* problem (CAP) in cellular network in terms of a *generalized vertex coloring* problem. Describe a method by which a CAP with *non-homogeneous* demand can be solved by solving a sequence of smaller subproblems where each subproblem has *homogeneous* demand. [3+7=10]
   (b) State the *perturbation-minimizing* frequency assignment problem (PMFAP). Explain with an example the *unforced assignment* (UA) and the *forced assignment with rearrangement* (FAR) operations used to solve the PMFAP. [3+(3+4)=10]

Indian Statistical Institute

Mid-Semester Examination (2018-2019)
M.Tech. (CS) II
Advanced Algorithms for Graph and Combinatorial Optimization Problems

Date: September 12, 2018 Maximum Marks: 60 Time: 2.5 hours

Answer as many as you can. The maximum you can score is 60.
Marks allotted to each question are indicated within parentheses near the right margin.

1. In an undirected connected graph G, a set of vertices C is called a *clique* if every two vertices of C are connected by an edge.
   (i) Prove that in the spanning tree resulting from DFS, all the vertices of a clique appear on one directed path.
   (ii) Do they necessarily appear consecutively on the path? Justify your answer.

   [5+5=10]

2. A directed graph $G = (V, E)$ is *singly connected* if $u - - - > v$ implies G has at most one simple path from u to v for all vertices $u, v \in V$. Give an efficient algorithm to determine whether or not a directed graph is singly connected. What is the worst case time complexity of your algorithm?

   [7+3=10]

3. Give an efficient algorithm to count the total number of directed paths in a directed acyclic graph. Analyze your algorithm.

   [6+4=10]

4. Present the steps of a version of Floyd-Warshall's algorithm for finding all-pairs shortest paths in a given directed weighted graph $G = (V, E, w)$ which requires $O(n^2)$ space.

   [8]

5. Decide whether you think the following statement is true or false. If it is true, give a short explanation. If it is false, give a counterexample.
   Let G be an arbitrary flow network, with a source s, a sink t, and a positive integer capacity c_e on every edge e. If f is a maximum s-t flow in G, then f saturates every edge out of s with flow (i.e., for all edges e out of s, we have $f(e) = c_e$).

   [7]

6. Recall the preflow-push algorithm for finding the maximum flow in a given network with n nodes and m arcs. Show that (i) the time to relabel all the nodes of the network is $O(nm)$, (ii) the number of saturating pushes is $O(nm)$.

[5+5=10]

7. Given a flow network $N = (V, E, c, s, t)$, sketch an algorithm to find the minimum flow from s to t. Argue briefly the correctness of your algorithm

[5]

8. Give the key ideas of Orlin's algorithm for finding the max-flow in a given network.

[5]

9. In a given undirected non-bipartite graph G and a matching M of its edges, show that while searching for an augmenting path from a node u of G with respect to the matching M, a blossom b is found. Then, if there is an augmenting path from u in the graph G/b with the blossom b shrunk in G, then there exists an augmenting path in G from u.

[5]

10. Trace the DFS based planarity testing algorithm on the bipartite graph $K_{3,4}$.

[10]

Indian Statistical Institute
M. Tech. (Computer Science)
Mid-Semestral Examination, 2nd Year, 2018-2019
Date: September 12, 2018
Subject: Natural Language Processing

NOTE:

Duration: 2 hr 30 min
Total Marks: 60 [Q1: 15, Q2: 7, Q3: 20, and Q4: 18]
Maximum Score: 50

---

1. Consider the following PCFG:

S $\longrightarrow$ SUBJ PRED	1.0
SUBJ $\longrightarrow$ NP	0.5
SUBJ $\longrightarrow$ NN	0.4
SUBJ $\longrightarrow$ PRN	0.1
PRED $\longrightarrow$ VB	0.4
PRED $\longrightarrow$ VB NP	0.6
NP $\longrightarrow$ DET NN	0.7
NP $\longrightarrow$ NP PP	0.3
PP $\longrightarrow$ PREP NP	1.0
PRN $\longrightarrow$ `i`	1.0
VB $\longrightarrow$ `like`	1.0
DET $\longrightarrow$ `this`	0.2
DET $\longrightarrow$ `the`	0.8
NN $\longrightarrow$ `make`	0.1
NN $\longrightarrow$ `shirt`	0.9
PREP $\longrightarrow$ `of`	1.0

(a)[**6 Marks**] Using the above PCFG, compute the probability of the following sentence: `i like the make of this shirt`

(b)[**6 Marks**] Use and n-gram language model to compute the probability of the above sentence in (a) if the available unigram and bigram statistics are as follows:

Unigram: `i`: 1×10^{-4}, `like`: 2×10^{-5}, `the`: 1×10^{-3}, `make`: 2×10^{-5}, `of`: 8×10^{-4}, `this`: 5×10^{-4}, and `shirt`: 1×10^{-5}.

Bigram: `i like`: 2×10^{-6}, `like the`: 1×10^{-6}, `of this`: 3×10^{-6}, and `this shirt`: 5×10^{-7}.

(c)[**3 Marks**] Compare the probabilities for the same sentence based on PCFG and n-gram statistics and reason behind the observed difference.

2. Prove that for an n-gram HMM-based POS tagging model with k distinct tags, the Viterbi decoding time of a t length observation is bounded by $O(k^n \times t)$.[**7 Marks**]

3. Consider the bi-gram HMM given in Figure 1. There are two tags N and V which can emit words. Two special tags 'start' and 'stop' cannot emit words. The vocabulary set = {'I', 'You', 'He'}. The word emission probabilities are given below.

   $p(\text{'}I\text{'}|N) = 0.2$, $p(\text{'}You\text{'}|N) = 0.4$, $p(\text{'}He\text{'}|N) = 0.4$
   $p(\text{'}I\text{'}|V) = 0.5$, $p(\text{'}You\text{'}|V) = 0.4$, $p(\text{'}He\text{'}|V) = 0.1$

   The observation sequence is 'He I He'.

   (a) Find the probability of the observation using backward probabilities. (**8 Marks**)

   (b) Assuming the values of all the parameters of the above HMM as initial values and the given observation, estimate $p(\text{'}He\text{'}|N)$ and $p(\text{'}He\text{'}|V)$ for the next iteration. (**12 Marks**)

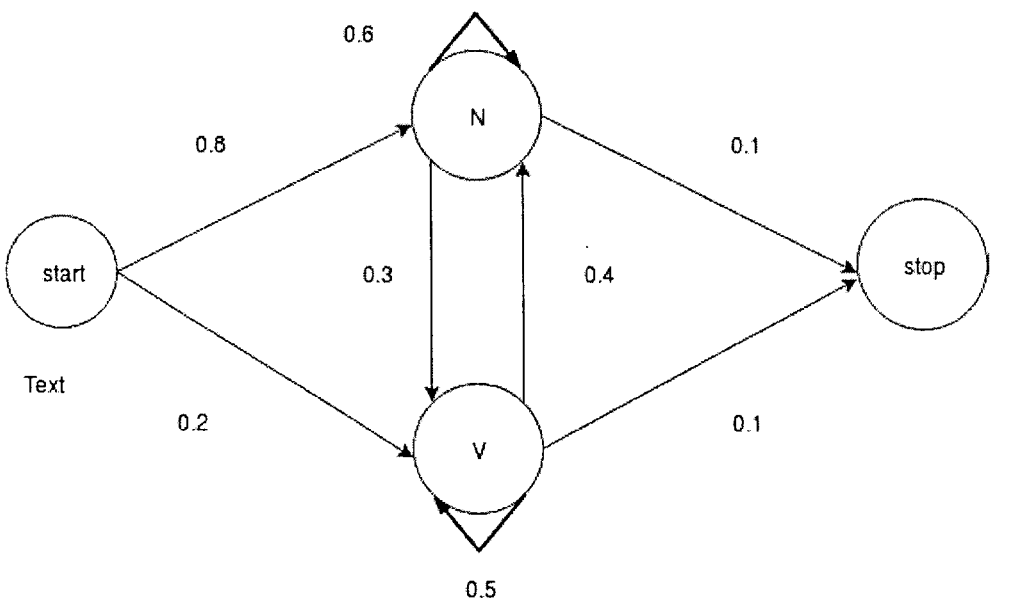


Figure 1: Automaton for the HMM.

4. A new model for statistical machine translation is defined as follows:

$p(f, a|e, m) = \prod_{i=1}^{m} t(f_i|e_{a_i}) \times q(a_i|a_{i-1}, l, m)$

Here f is a French sentence $f_1, \ldots, f_m$. a is the sequence of alignment variables $a_1, \ldots, a_m$ and e is an English sentence $e_1, \ldots, e_l$. t is the translation parameter and q is the alignment parameter. We assume that a_0 is defined to be 0. Note that in contrast to the IBM Model 2, the alignment parameters are modified to be conditioned upon the previous alignment variable. That is, alignment of i^{th} French word depends on the alignment of $(i-1)^{th}$ French word.

Given the training data (f^k, e^k) for $k = 1$ to n where lengths of f^k and e^k are m_k and l_k respectively.

(a) Define the *count*() variables for parameter estimation of the model. **(10 Marks)**

(b) Write down the model's parameter estimation algorithm for fully observed data. **(8 Marks)**

INDIAN STATISTICAL INSTITUTE

M.TECH. (CS) - YEAR II

2018 – 2019

LOGIC FOR COMPUTER SCIENCE

DATE : 13/09/2018

Time: 3 Hours MID-SEMESTER EXAMINATION **Marks: 40**

Answer Question 1 and any 4 from the rest.

1. Let $\mathcal{L}$ be a propositional language whose formulas φ are given as follows:

   $$p \mid \neg\varphi \mid \varphi \to \psi$$

   where $p \in \mathcal{P}$, a countable set of propositions. Let $\mathcal{V} : \mathcal{P} \to \{0, 1\}$ be a valuation function. The truth definition of the formulas is given as follows:

   $\mathcal{V} \models p$ iff $\mathcal{V}(p) = 1$;
   $\mathcal{V} \models \neg\varphi$ iff $\mathcal{V} \not\models \varphi$;
   $\mathcal{V} \models \varphi \to \psi$ iff $\mathcal{V} \not\models \varphi$ or $\mathcal{V} \models \psi$.

   Consider the following axiom scheme:

   Axiom 1: $\varphi \to (\psi \to \varphi)$
   Axiom 2: $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$
   Axiom 3: $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$

   and the rule:

   $$\frac{\varphi \qquad \varphi \to \psi}{\psi}$$

   Let Γ be a set of formulas in the given language, and φ is a formula in the language. Define $\boldsymbol{\Gamma \vdash \varphi}$ if there is a sequence of formulas $\varphi_1, \ldots, \varphi_n$ such that φ_n is φ and each φ_i is either an instance of an axiom (mentioned above), or a member of Γ or obtained by the rule mentioned above. Define $\boldsymbol{\Gamma \models \varphi}$ if for every valuation $\mathcal{V}$, $\mathcal{V} \models \varphi$, whenever $\mathcal{V} \models \gamma$ for all $\gamma \in \Gamma$. Define **Γ to be consistent** if there is a formula φ, such that $\Gamma \vdash \varphi$ and $\Gamma \nvdash \varphi$. Define **Γ to be maximally consistent** if (i) Γ is consistent, and (ii) for every formula $\varphi \notin \Gamma$, $\Gamma \cup \{\varphi\}$ is not consistent.

   Show the following:

   (a) Every consistent set Γ can be extended to a maximal consistent set.
   (b) If Γ is a maximal consistent set, then there exists a valuation $\mathcal{V}$, such that $\mathcal{V} \models \gamma$ for all $\gamma \in \Gamma$. [20]

2. Let $\mathcal{A}$ be a non-empty set. A binary relation $<$ on $\mathcal{A}$ is said to be linearly ordered if $<$ is irreflexive, transitive and satisfies the law of trichotomy. The relation $<$ is dense if for any $x, y \in \mathcal{A}$, with $x < y$, there is some $z \in \mathcal{A}$ such that $x < z < y$. Find a suitable first order language and give axioms in it for a dense linear order without bounds. [5]

3. Consider $L_<$, a first order language with equality and a binary predicate symbol $<$, and the structure $(\mathbb{Q}, <_{\mathbb{Q}})$, where $\mathbb{Q}$ denotes the set of rational numbers, and $<_{\mathbb{Q}}$ is the usual strict ordering in $\mathbb{Q}$. Which of the following sentences are true in this structure ?

   (a) $\forall x \exists y (x < y)$.
   (b) $\exists x \forall y (x < y \to x = y)$.
   (c) $\forall x \forall y \forall z (x < y \to (y < z \to x < z))$. [5]

4. Consider a first order language with equality and a 2-place predicate symbol P. Construct the following:

   (a) a sentence φ in this language such that φ is true in the structure $(\mathbb{R}, <)$ but not in the structure $(\mathbb{N}, <)$.

P.T.O

(b) a sentence ψ in this language such that ψ is true in both the structures.

Here $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers and $<$ is the standard strict ordering on these sets. [5]

5. Check whether the following formulas are true in every first order structure:

   (a) $\forall x Px \to \exists x Px$

   (b) $\exists x \forall y Pxy \to \forall y \exists x Pxy$ [5]

6. Give examples of structures and assignments in those structures under which the following formulas are not satisfied.

   (a) $\exists x Px \to \forall x Px$

   (b) $\forall y \exists x Pxy \to \exists x \forall y Pxy$ [5]

7. Let $\mathcal{R}$ be the structure $(\mathbb{R}; +_{\mathbb{R}}, \cdot_{\mathbb{R}})$, $\mathbb{R}$ being the set of real numbers, $+_{\mathbb{R}}$ and $\cdot_{\mathbb{R}}$ correspond to the usual addition and multiplication, respectively. Let $\mathcal{L}$ be a first order language with equality, two binary function symbols $+$ and $\cdot$, interpreted as $+_{\mathbb{R}}$ and $\cdot_{\mathbb{R}}$, respectively in the structure $\mathcal{R}$.

   (a) Give a formula in $\mathcal{L}$ that defines the set $\{x \in \mathbb{R} : x \geq 0\}$ in the structure $\mathcal{R}$.

   (b) Give a formula in $\mathcal{L}$ that defines the set $\{0\}$ in the structure $\mathcal{R}$. [5]

Indian Statistical Institute

M. Tech (Computer Science) Second Year (Semester I)

Mid-Semester Examination - 2018

Pattern Recognition and Image Processing

Time: **2 hours** Total Marks: 70 Maximum Marks: **60**

Date: 13/09/2018

1. Short Questions 2x5=10
   a. Explain accuracy of a classifier.
   b. How would you handle an imbalanced dataset?
   c. What is dimensionality reduction?
   d. How can you ensure that a given model is not overfitting the data?
   e. Suggest two methods of handling missing values in the data.

2. Suppose you are asked to predict whether a political party will win the majority in each of five zones going for poll. 4+3+3=10
   a. What are the inputs and features you would like to use to determine the winning party if you have no previous data?
   b. Is this a classification problem or a clustering problem and why?
   c. How will you handle the problem if you only have data for the previous years?

3. Discuss in details **one** approach for feature selection, with illustration. 10

4. Consider various distance measures. 4+(2+2+2)=10
   a. What properties should a metric **d** on a vector space **X** satisfy?
   b. Two objects are represented by the tuples (32, 1, 22, 6) and (20, 0, 12, 8):
      i. Compute the Euclidean distance between the two objects.
      ii. Compute the Manhattan distance between the two objects.
      iii. Compute the Minkowski distance between the two objects, using power n=3.

5. You are given feature vectors $\mathbf{x}_1$ from class ω_1 and feature vectors $\mathbf{x}_2$ from class ω_2. 2+3+4+1=10
   The training set consists of the following points:
   Points from Class 1: {(11, 11), (13, 11), (8, 10), (9, 9), (7, 7), (7, 5), (15, 3)}
   Points from Class 2: {(7, 11), (15, 9), (15, 7), (13, 5), (14, 4), (9, 3), (11, 3)}
   a. Draw a scatter plot.
   b. Are the two classes linearly separable? Explain.
   c. In the scatter plot, also sketch the decision boundary that you would get using a nearest neighbour classifier (assume k=1 for kNN).
   d. Classify the sample (6, 11) using the classifier thus obtained.

6. Given the points A = (1, 2), B = (2, 2), C = (2, 1), D = (-1, 4), E = (-2, -1), F = (-1, -1) 8+2=10
   a. Starting from initial cluster centres for Cluster1: A and Cluster2: D, run the K-means clustering algorithm and report the final clusters. Use L1 distance as the distance between points.
   b. Draw the points on a 2-D grid and check if the clusters make sense.

7. For the minimum distance classifier 4+2+4=10
   a. Show that the classification boundary is a perpendicular bisector of the line joining the two means for a binary classification problem.
   b. How will the decision boundary look for a three class classification problem?
   c. Show that the statement in (a) holds for any D dimensional problem.

Indian Statistical Institute
Mid-Semester Examination (2018)
M.Tech.(CS) II Year
Cognitive Science

Date: 14.09.2018 **Full Marks: 75** **Duration – 3 hours**

Answer as many questions as you like, but you can at most score 75.

1. (a)Complex systems, including biological ones are often understood in terms of the activities of simpler building blocks or agents. How a child's brain or any cognitive system at work, can in general, be thus explained algorithmically? (b) Can you tell a few words about any such work done at the MIT AI Lab? (c) How can Machine Intelligence be explained in the light of (a)? (d)Give an example of neuropsychological test that can assess intelligence of children? What is the main contribution of William Stern in this domain? [6+2+4+2+1=15]

2. (a) What conclusion can you draw about learning from Piaget's experiments? (b) Can we extend Piaget's experiment to the adult brain using, for instance, the following example: Fill up the blank with the appropriate word according to you: This is a story of an incident that occurred many, many years ago in a ..... (*nearby/far away*) land? Justify. Is Machine Learning explainable in the same way? Justify your answer again. [6+4+2=12]

3. (a) What is the basic unit of the central nervous system, and who would you give credit for identifying the same? (b)What are the two main components of its membrane potential (V)? (c)Write down these two components and form the differential equation for the flux of any ion across the membrane. (d) Solve the equation and arrive at the generalized expression for membrane potential for any ion. (e) What is this potential called? (f) If the ratio of the molar concentrations of Na^+ ion outside and inside the membrane is 0.1, then calculate V_{Na}. [3+2+4+4+1+4=18]

4. Find the correct alternative from the options given below.

(I) The purpose of neuropsychological assessment is to:

a) Identify cognitive impairment

b) Evaluate preserved cognitive ability

c) Both the above

d) None of the above

(II) A comprehensive series of cognitive tests that lead to one general score is called:

a) Battery

b) Scale

c) Compendium

d) None of the above

(III) A person's level of cognitive ability prior to brain injury is called:

a) Pre-injury intelligence

b) Premorbid intelligence

c) Fluid intelligence

d) None of the above

(IV) Find the appropriate one for any good neuropsychological test:

a) Give consistent scores for an individual after repeated observations ; males get advantages over females; make accurate prediction of performance

b) Give inconsistent scores for an individual after repeated observations; no gender biasness; make inaccurate prediction of performance

c) Give consistent scores for an individual after repeated observations; no gender biasness; make accurate prediction of performance

(V) If you want to measure the IQ of a girl having age of 10 years, which part of Raven's Progressive Matrices will be appropriate for that:

a) Advanced

b) Standard

c) Coloured

(VI) You have to assess the IQ of two different types of children having same age. But one is attending school and another one is ill-literate. In this situation which of the flowing test you will select to assess their IQ:

a) Wechsler Intelligence Scale for Children

b) Raven's Progressive Matrices

c) Stanford-Binet Test

Explain your answer.

(VII) Fill in the gap: Watson and Crick are to cell replication what ..... is/are to cell machines:

a) Kurt Godel

b) John von Neumann

c) McCulloch and Pitts

d) Freud and Piaget

(VIII) You are an elementary school psychologist. As you look through the results of recent IQ tests given to your students, you take note of several students who have scored well on the non-verbal subscale and poorly on the verbal subscale. Based on these results, you decide to bring these students into your office and

a) Administer a battery of tests to assess the possibility of learning disabilities.

b) Interview them about any difficulties they may be having at home.

c) Administer a battery of personality tests to create a psychological profile of each student's ability to adapt to stressful situations.

[1+1+1+1+1+(1+2)+1 +1=10]

5. (a)What is the difference between visual perception and visual cognition? (b) How do you mathematically define eccentricity? Hence describe the importance of retinal eccentricity in physical filling-in for gap in vision. (c) Describe the activities of an ON Center-OFF Surround Ganglion cell for the illumination conditions mentioned below:
   (I) Center and Surround both are illuminated fully.
   (II) There is light on the Surround,and Center is illuminated partly.

   [3+(2+3)+(2+2)=12]

6. (a) In a study, Inferior temporal gyrus and anterior of the temporo-occipital lobe of the brain were found to be active on presentation of visual numerals to a subject. This indicates that the region is a specialized region involved in processing of visual numerals in those subjects. Does it indicate that this region evolved for this ability via natural selection? (b) If yes, list a similar organ or brain region that evolved this way? Else list a brain region or organ that initially evolved for something else and later underwent modification. (c)

Could genes affecting development of the example you presented, and the regions of processing numbers have underwent modification? [2+3+3=8]

7. (a) Human beings can learn and master new skills in their lifetime. Does it mean that human beings can master every skill or everyone can master the skills with same proficiency? (b) If not, why? (c) Do songbirds come with songs ingrained in their brains? (d) How can the singing abilities of birds evolve, and do songs help in their survival? (e) Many similar sounding words exist in Indian languages and European languages, e.g. *name* & *naam* or *mom* and *ma*. Do you think that such similarity is the result of some intrinsic property of human brains that had its impact on evolution of words and languages? Justify your answer. [2+3+1+3+3=12]

M.Tech (CS) – II[nd] Year

Mid-Semester Exam - 2018

Artificial Intelligence

14.09.2018

Full Marks – 100 Duration – 3 hours

Answer any five.

1. Derive Skolem standard form and clausal form of the following expression:
$$(\forall x)((\exists y)\, P(x,y) \wedge \sim Q(y,x)) \vee ((\forall y)(\exists z)(R(x,y,z) \wedge \sim Q(y,z)))$$
(20)

2. Suppose we can perform the following chemical reactions:
$$MgO + H_2 \rightarrow Mg + H_2O$$
$$C + O_2 \rightarrow CO_2$$
$$CO_2 + H_2O \rightarrow H_2CO_3$$
Suppose we have some quantities of MgO, H_2, O_2 and C. Using resolution principle show that we can make H_2CO_3. (20)

3. a. Define Herbrand universe $H_0, H_1, \ldots\ldots, H_\infty$. (5)
b. Let, $S = \{P(f(x), Q, g(f(x), b))\}$.
Find
   1. H_0 and H_1
   2. all ground instances of S over H_0.
   All ground instances of S over H_1. (15)

4. Let us consider the two universes volume and mass in standard units, that is, m^3 and kg respectively and represent the following two crisp sets;
$$X = Volume = \{10, 15, 20, 25\} \text{ and}$$
$$Y = Mass = \{40, 60, 80, 100\}.$$
Let us consider the following conditional statement;
IF x is small THEN y is heavy.
Let x is small $= \{{}^{1}/_{10}, {}^{0.9}/_{15}, {}^{0.5}/_{20}, {}^{0.1}/_{25}\}$
and y is heavy $= \{{}^{0.3}/_{40}, {}^{0.6}/_{60}, {}^{0.8}/_{80}, {}^{1}/_{100}\}$.
Convert the above IF-THEN statement into fuzzy relation R using Mamdani's 'min' operator. (20)

P.T.O.

5. a. What is Vanishing Gradient Problem? Explain.
   b. What are the remedial measures that can be taken to prevent it?
   c. In relation to the above question, how does the remedial measures solve the Vanishing Gradient Problem?
   d. What are the issues associated with using ReLU?
   e. What are the alternatives to ReLU, that can be used to avert those issues?
   f. How is non-linearity modelled using ReLU?
   g. What are the problems associated with zero initialization in Neural Networks?
   h. What will happen, if all the weights, in the above Neural Network, is initialized with 1?

   (4+1+3+2+2+3+3+2 = 20)

6. a. What is Bias-Variance Tradeoff?
   b. Explain briefly how L2 regularization and Dropout reduces overfitting.
   c. What are the effects on the weights of a neural network, if the inputs are not normalized?
   d. What are the effects on the Optimization process, if the inputs are not normalized?
   e. What is the advantage of Gradient Descent with Momentum over Vanilla Gradient Descent?
   f. Draw the computation graph for a 2-layer Feedforward Neural Network. Choose any number of inputs and nodes in the hidden layers.
   g. Write a Python code, using TensorFlow, to implement the forward propagation steps of the above neural network.

   (3+(1.5+1.5)+2+2+3+2+5 = 20)

7. a. Why is Convolutional Neural Network preferred over Feedforward Neural Network, for learning image representations/features?
   b. What is degradation problem? How was this problem solved using shortcut connections?
   c. Any shallower network is a subspace of a deeper network. True or False.
   d. Mention a disadvantage of VGG Net.
   e. What is the utility of 1x1 convolutions in GoogLeNet architecture?
   f. Mention a disadvantage of R-CNN.
   g. What changes were done in Fast R-CNN?
   h. In YOLO, a network pre-trained on ImageNet dataset has been used, and then fine-tuned for object detection. Can YOLO be used to make a face detector? Answer in Yes or No.
   i. If the answer to the above question is Yes, then briefly explain how can it be implemented? If No, then justify.
   j. How can the YOLO architecture be trained to recognize faces?

   (3+3+1+1+3+1+1+1+3+3 = 20)

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2018

Date : 14/09/2018

Subject Name : **Cryptology** Course Name : M.Tech. (CS) II yr

Max Score: 40 Duration: 150 Mins

Note: Attempt all questions. Marks are given in brackets. Total score is 46. But maximum you can score is 40. Use separate answer book for Q6-Q8.

1 [2] Apply Babbage-Kasiski test to estimate the key length of the Vigenère cipher used to produce the following ciphertext.

```
DSZOS PXSRD SZOXF KXGCX FOUSO WRSSL
```

2 [5] Let $\mathcal{E} = (e, d)$ be a semantically secure encryption scheme defined over $(\mathcal{K}, \{0,1\}^n, \{0,1\}^n)$. Let $\mathcal{E}' = (e', d')$ be an encryption scheme defined over $(\mathcal{K}, \{0,1\}^{3n}, \{0,1\}^{3n})$ where

$$e'_K(m_1, m_2, m_3) = (e_K(0) \oplus m_1, e_K(1) \oplus m_2, e_K(0) \oplus e_K(1) \oplus m_3).$$

Show that $\mathcal{E}'$ is not semantically secure.

3 [2+4=6] Let X and Y be two random variables over a finite set Ω. Define the statistical distance $\Delta(\mathsf{X};\mathsf{Y})$. Let $\mathsf{Ind}(\Omega)$ denote the set of all functions from Ω to $\{0,1\}$. Prove that

$$\Delta(\mathsf{X};\mathsf{Y}) = \max_{f \in \mathsf{Ind}(\Omega)} \mathbf{Ex}[f(\mathsf{X})] - \mathbf{Ex}[f(\mathsf{Y})].$$

4 [2+4=6] Define a (t, ϵ)-Pseudorandom Bit Generator (PRBG). For any $\epsilon < \frac{1}{2}$, show that there does not exist an (∞, ϵ)-PRBG.

5 [6] Let $n, m \in \mathbb{N}$, $m > n$, and $f : \{0,1\}^n \to \{0,1\}^{n+1}$. Construct a new function $g : \{0,1\}^n \to \{0,1\}^{m+n}$ using the Bloom-Mikali sequential composition of f. Prove that if f is a secure PRBG then g is also a secure PRBG.

6 [6] Let $\mathcal{E} = (KG, E, D)$ be a public key encryption scheme with a message space G (which is an additive group) and a ciphertext space C. The PKE $\mathcal{E}$ is said to be additive if there is a combiner function $f : C \times C \to C$ such that $D(c) + D(c') = D(f(c, c'))$ for all valid ciphertext c, c'. Give an example of additive PKE and justify.

P.T.O.

7 [6] Let A be an efficient algorithm which can solve discrete log problem with probability $1/2$ for a random challenge (i.e. finding x from g^x where x is chosen randomly). Construct a similar efficient algorithm B which can solve discrete log problem with probability at least $1/2$ in the worst case (i.e. for all x, B should find x from g^x with probability at least $1/2$).

8 [2+2+5=9] Consider the following modified RSA signature algorithm. For same RSA modulus $n = pq$, it generates two pairs (e_1, d_1) and (e_2, d_2) where (d_1, d_2) is the signing key and (n, e_1, e_2) verification key. Signing algorithm takes a message $(m_1, m_2) \in \mathbb{Z}_n^*$ and computes signature $\sigma = m_1^{d_1} m_2^{d_2} \mod n$.

1. Write down the verification algorithm.
2. Define universal forgery game.
3. Construct a universal forgery for the above signature algorithm which makes only one signing query.

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2018
Course Name: M.Tech. In Computer Science
Subject Name: Computer Architecture

Date: 15.09.2018 *Maximum Marks: 40* *Duration: 2 hours*

Answer all questions

1. For each of the following, mention if the concept is a part of the ISA or the microarchitecture: [3]

   - Number of threads in fine-grained multi-threading
   - Number of DRAM banks
   - Vectored Interrupts
   - Number of entries in reservation stations
   - Number of entries in reorder buffer
   - Number of entries in the architectural register file

2. Among the following sets of two concepts, circle the one that makes programmer's task harder and **at the same time** micro-architect's (or hardware designer's) task easier. Interpret programmer broadly as we did in class, e.g. a compiler writer is also a programmer [4]

   - Reduced instruction set versus complex instruction set
   - VLIW versus superscalar execution
   - More addressing modes versus less addressing modes
   - Unaligned access support versus alignment requirements in hardware

3. Assume you have a machine with a **4-entry** return address stack. A code executing on this machine has **six** levels of nested function calls, each of which end with an appropriate return instruction. What is the return address prediction accuracy for this code? [2]

4. Consider the following high-level programs:

   Program A:
   ```
   int8 a[150] // a is allocated in memory
   for (i = 0; i < 150; i++) {
         a[i] = 5;
   }
   ```

   Program B:
   ```
   int *p; // *p is allocated in memory
   *p = 150;
   ```

   Program C:
   ```
   int **p; // *p and **p are allocated in memory
   **p = 150;
   ```

   Assume in Program A, a register contains the address of the start of the array, and in Programs B and C, a register contains the value of p.

   For each of the above three programs, which addressing mode would lead to the minimum number of instructions? Explain your answer briefly. [2+2+2]

5. You are designing an ISA that uses delayed branch instructions. You are trying to decide how many instructions to place into the branch delay slot. How many branch delay slots would you need for the following different implementations? Explain your reason briefly. [2+3]

   - An in-order processor where conditional branches resolve during the 4th stage
   - An out-of-order processor with 2 Functional units, with a unified 16-entry reservation station with a 15-stage pipeline with conditional branches being resolved at the 2nd cycle of branch execution

6. You designed a microprocessor. It came back from the fab with an error: one of the bits is stuck. We call the bit a stuck-at-0 fault if the bit is always 0 (i.e., you cannot store a 1 in it). We call the bit a stuck-at-1 fault if the bit is always 1 (you cannot store a 0 in it).

   Consider each of the structures below independently. Assume the structure contains a stuck at 0 or 1 fault. Does the fault affect the correctness of the chip? Does the fault affect performance? Explain.
   - A bit in the register scoreboard
   - A bit in the "predicted next instruction address" supplied by the branch predictor
   - A valid bit in the reorder buffer

   Note: For each structure, consider separately stuck-at-0 and stuck-at-1 faults. Assume no error detection and correction mechanisms are present. [2+2+2]

7.
   a) Remember that the history buffer is a structure that enables precise exceptions in a pipelined machine. Briefly describe how the history buffer works during the execution of a register-to-register ADD instruction (ADD RD ← RS1 + RS2). [2]
   b) Assume we would like to use the exact same solution (history buffer) for executing a store instruction to memory. Why is this difficult to do? [2]
   c) Briefly describe one solution to handling store instructions in conjunction with a history buffer [2]

8. Given the following code:

```
MUL  R3, R1, R2
ADD  R5, R4, R3
ADD  R6, R4, R1
MUL  R7, R8, R9
ADD  R4, R3, R7
MUL  R10, R5, R6
```

   Note: Each instruction is specified with the destination register **rst** as the first operand.

Calculate the number of cycles it takes to execute the given code on the following models:
- A non-pipelined machine
- A pipelined machine with score-boarding and five adders and five multipliers without data forwarding
- A pipelined machine with score-boarding and five adders and five multipliers with data forwarding.
- A pipelined machine with score-boarding and one adder and one multiplier without data forwarding
- A pipelined machine with score-boarding and one adder and one multiplier with data forwarding

Note: For all machine models, use the basic instruction cycle as follows:
- Fetch (1 clock cycle)
- Decode (1 clock cycle)
- Execute (MUL takes 6, ADD takes 4 clock cycles). The multiplier and the adder are not pipelined.
- Write-back (one clock cycle)

Do not forget to list any assumptions you make about the pipeline structure (e.g. how is data forwarding done between pipeline stages) [2+2+2+2]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination:(2018-2019)

M.TECH (CS) II YEAR

Subject Name: Quantum Information Processing and Quantum Computation

Maximum Marks: 30 Duration: 2 hours Date: 15/09/18

Answer any three of the following four questions

1. a) Consider the following linear operator A acting on a two dimensional Hilbert space.;

$$A = aI + n.\sigma$$

where $n.\sigma = \sum n_i\sigma_i$ $(i = x, y, z)$, n is a vector in R^3, σ's are Pauli matrices and I is identity operator and a is real.
i) Under what condition A is a density operator .
ii) Under what condition A is a projection operator.
b) Show that for every the density operator D with $D^2 = D$; there exists a unit vector $|\psi\rangle$, such that

$$D = |\psi\rangle\langle\psi|$$

c) Let the initial density matrix of a qubit is $\frac{1}{2}(I+\frac{1}{3}\sigma_x+\frac{1}{2}\sigma_y)$. If spin measurement is performed along z-axis, what is the probability for the spin up result?
(2+2)+3+3

2. a) Let there is a cloning machine that can clone the following orthogonal states $|0\rangle$ and $|1\rangle$. Show that the machine will not be able to clone the following state $|\psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$.

b) Consider a Swap operator U_s which acts in the following way;

$$U_s|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$$

for all possible states $|\psi\rangle$, $|\phi\rangle$. Then show that U_s can not be written as

$$U_s = U_1 \otimes U_2$$

where U_1 and U_2 are acting on particle 1 and particle 2 respectively.

c) Describe the realization of the swap operation (gate) by using C-Not gates. $4+3+3$

3. a) Let Alice and Bob share the following state;

$$|\psi>_{AB}= \frac{1}{\sqrt{3}}(|0\rangle_A \otimes |0\rangle_B + \sqrt{\frac{2}{3}}|1\rangle_A \otimes |1\rangle_B)$$

where $|0\rangle$ and $|1\rangle$ are eigen states of σ_z.
(i) Show that the state can not be written in product form.
ii) Find the density matrix of the subsystem on Bob's side.
(ii) Derive the probability of conclusive teleportation of an unknown state by using this state. How many bits will be required for conclusive teleportation? $2+2+6$

4 a) Argue why by sending an isolated single qubit, no more than 1 bit of information can be communicated though there are infinitely many possible states.
Show how by sending a single qubit which is in a maximally entangled state with another qubit in the possession of the receiver, 2 bits of communication can be made.
b) Consider a function $f : \{0,1\}^n \rightarrow \{0,1\}$, where the function f is either constant or balanced ($f(x) = 0$ for half of the possible input values).
Discuss how difficult it is to determine whether the function is constant or not in classical world.
Describe the quantum algorithm by which the function can be shown to be either constant or balanced by just one implementation of a quantum circuit. (1+4)+5

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2018-2019

M. Tech. (CS) II year

Data Mining and Knowledge Discovery

Date: 10.09.2018 **Maximum Marks: 70** **Duration: 2 hours 30 mins**

[Answer as much as you can]

1. (i) What do you understand by the "scalability" of an algorithm?
   (ii) How is a decision tree different, as a classifier, with respect to neural networks?
   (iii) What is the perceptron learning rule? When does it fail?
   (iv) What is overfitting? How is it different from underfitting?
   (v) Outline some stopping criteria for a classifier. [4+4+5+5+4=22]

2. (i) Explain algorithm SPRINT, pointing out its merits and demerits in the context of data mining.
   (ii) Elaborate on some performance measures for evaluating classifiers.
   (iii) What is ROC? How is it employed to compare classifier models? [10+6+5=21]

3. (i) Describe page rank algorithm.
   (ii) Elaborate on Q-learning.
   (iii)Explain ARMA and ARIMA
   (iv)What is one shot learning, in the context of Siamese NNs?
   (v)How is SVM used as a classifier? [5+5+5+5+5=25]

4. Figure shows a very small neural network: it has one input unit, one hidden unit (logistic), and one output unit (linear). Let us consider one training case, for which the input value is 1 (as shown in the diagram), and the target output value is 1. We are using the standard squared error loss function:

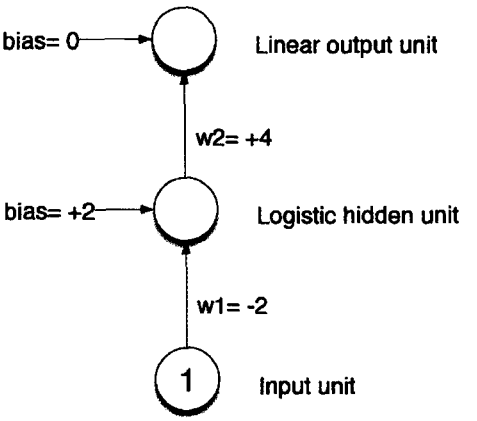
   $E = (t - y)^2 / 2$

(ii) What is the output of the hidden unit and the output unit, for this training case?

(iii)What is the loss, for this training case?

(iv)What is the derivative of the loss w.r.t. w2, for this training case?

(v) What is the derivative of the loss w.r.t. w1, for this training case?



[2+1+2+2=7]

5.

(i) We know that the momentum hyper-parameter enables us to temporarily smoothen out the gradient samples obtained through stochastic gradient descent. What is/are the main advantage (s) of this technique? What effect does increasing the momentum hyper-parameter (in other words, increasing the weight of previously sampled gradients) have on the model's capacity?

(ii) Hinton indicated about " dividing the learning rate for a given weight by a running average of the magnitude of recent gradients for that weights".

a. What is the intuition behind this method?

b. What problems does it help to solve?

(iii) In the case of stochastic gradient descent method a faster convergence has been observed if the order in which the *mini-batches are visited is changed for each epoch.* Do you have a plausible explanation for that?

(iv)Let $f(x)$ be an activation function. Show that the inverse function $x = f^{-1}(x)$ increases with f if $\frac{df}{dx} > 0$.

[4+(2+2)+3+4=15]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M.Tech.(CS) – II Year, 2018-2019 (Semester – 1)

Optimization Techniques

Date: 13.11.2018 Maximum Marks: 100 Duration: 4 hours

Note: The question paper is of 150 marks. Answer as much as you can, but the maximum you can score is 40 from Group-A and 60 from Group-B.

Notations: Vectors would be written in small letters with boldface, e.g. $\mathbf{b}$; matrices would be written in capital letters, e.g., A. Transpose of A would be denoted by A^T and transpose of $\mathbf{b}$ would be denoted by $\mathbf{b}^T$. Whenever we say that, $\mathcal{P}$ is a linear program, we mean $\mathcal{P}$ is of the form

$$\begin{array}{lrcl} \text{Maximize} & \mathbf{c}^T\mathbf{x} & & \\ \text{subject to} & A\mathbf{x} & \leq & \mathbf{b} \\ & \mathbf{x} & \geq & \mathbf{0} \end{array}$$

and $\mathcal{P}_{eq}$ will denote a linear program of the form

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

Let A and B be two $n \times n$ matrices, then we write $A \succeq B$ (and $A \succ B$) if $A - B$ is positive semi-definite matrix (and $A - B$ is positive definite matrix).

Let A be a $m \times n$ matrix and $b \in \mathbb{R}^m$, then $b \in \mathcal{R}(A)$ means there exists $x_b \in \mathbb{R}^n$ such that $Ax_b = b$.

Group-A

(AQ1) Consider a convex polygon P. The problem is to compute the maximum area circle that can lie properly inside (no part of the circle outside P) the polygon P. Thus, the objective is to compute the coordinates of its center (α, β), and the radius r. Formulate this problem as a linear programming problem. [10]

(AQ2) Consider the integer programming problem – $maximize\ c^T x$, subject to $Ax \leq b$, $x \geq 0$, the elements of x are integer valued.

where x is the vector of n variables; A is the coefficient matrix of the linear constraints, b is the requirement vector, and c is the cost vector. The entries of A, b, and c are all positive integers.

Let LP be the linear program obtained by relaxing the constraint that x be integer. Call the solution of the ILP as x_0 and the solution of LP as x_1.

Show that in the ILP (i) $x^* = \lfloor x_1 \rfloor$ is a feasible solution, and (ii) $|c^T x_0 - c^T x^*| \leq \sum_{i=1}^{n} c_i$.

[5+5=10]

(AQ3) (a) Show that, an integer matrix A with entries $a_{ij} \in \{-1, 0, +1\}$ is a *totally unimodular matrix* if no more than two non-zero entries appear in each column, and if the rows of A can be partitioned into two sets I_1 and I_2 such that

(i) if a column has two non-zero entries with same sign, their rows are in different sets, and
(ii) if a column has two non-zero entries of different sign, their rows are in the same set.

(b) Let A be a matrix with entries $a_{ij} \in \{0, 1\}$. Show that, if in each column of A, all the 1's appear in consecutive row positions, then A is a *totally unimodular matrix.* [5+5=10]

(AQ4) Consider the following mathematical program in $\mathbb{R}^n$:

$$\begin{array}{ll} \text{Minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \;\; i \in \{1, \ldots, m\} \\ & h_j(x) = 0, \;\; j \in \{1, \ldots, p\} \end{array}$$

Assume that $f_0, \ldots, f_m, h_1, \ldots, h_p$ be differentiable functions.

Let x^* and (λ^*, μ^*) (where λ^* and μ^* are the Lagrangian multiplier vectors for the inequality and equality constraints respectively) be any primal and dual optimal solutions to the above mathematical program with zero duality gap.

(a) State and prove the necessary KKT optimality conditions in terms of x^* and (λ^*, μ^*).

(b) State and prove the conditions under which KKT conditions are sufficient.

[5+5=10]

(AQ5) Consider the following mathematical program in $\mathbb{R}^n$:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

where f is a differentiable convex function and $A \in \mathbb{R}^{m \times n}$. Prove that, a point $x^* \in \mathbb{R}^n$ is optimal for the above mathematical program iff x^* is feasible and there exists $\mu^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) = A^T \mu^*.$$

[10]

(AQ6) Let p^* be the optimal value of the following mathematical program **(P1)** in $\mathbb{R}^n$:

$$\begin{array}{ll} \text{Minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

where $A \in \mathbb{R}^{m \times n}$ and let d^* be the following dual mathematical program **(P2)** in $\mathbb{R}^m$:

$$\begin{array}{ll} \text{Maximize} & -b^T y \\ \text{subject to} & A^T y + c \geq 0. \end{array}$$

(a) Using Lagrangian Multipliers method show that **(P2)** is the dual mathematical program to **(P1)**.

(b) Show that strong duality holds between **(P1)** and **(P2)**, i.e. $p^* = d^*$, provided at least one of the programs is feasible.

[5+5=10]

Group-B

(BQ1) Consider the following quadratic programming problem:

$$\begin{array}{ll} \text{Minimize} & C^TX + \frac{1}{2}X^THX, \\ \text{subject to} & AX \leq B \\ & X \geq 0, \end{array}$$

where $X, C \in I\!R^n$, $H \in I\!R^{n\times n}$, $A \in I\!R^{m\times n}$ $(m \leq n)$, and $B \in I\!R^m$.

Describe Wolfe's method to solve this problem.

[15]

(BQ2) Let $n \geq 2$, $S(0,1) = \{\mathbf{x} \in I\!R^n | \mathbf{x}^T\mathbf{x} \leq 1\}$ and $E = \{\mathbf{x} \in I\!R^n | (\mathbf{x}-\mathbf{t})^T B^{-1}(\mathbf{x}-\mathbf{t}) \leq 1\}$, where

$$\mathbf{t} = \begin{bmatrix} -\frac{1}{n+1} \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}_{n\times 1} \quad \text{and} \quad B = \begin{bmatrix} \frac{n^2}{(n+1)^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{n^2}{n^2-1} & 0 & \cdots & 0 \\ 0 & 0 & \frac{n^2}{n^2-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \frac{n^2}{n^2-1} \end{bmatrix}_{n\times n}$$

Show that

(i) B is positive definite and so E is an ellipsoid,

(ii) The hemisphere $HS = \{\mathbf{x} | \mathbf{x}^T\mathbf{x} \leq 1 \text{ and } x_1 < 0\} \subseteq E$.

(iii) $\frac{\text{volume}(E)}{\text{volume}(S(0,1))} < 2^{-\frac{1}{2(n+1)}}$.

[4+6+5=15]

(BQ3) In the context of solving a linear programming problem P (as defined in the preamble) using interior point method, define a barrier function, and the central path method.

Now, formulate the problem of approximately optimizing of the barrier function with respect to the given constraints using the central path method as a linear programming problem, and an idea of solving the given LP problem P.

[(2+2)+(8+3)=15]

(BQ4) Let $P, Q \subseteq \mathbb{R}^n$ be convex sets and let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Suppose that x^* is an optimum solution to $\min\{f(x) \mid x \in P \cap Q\}$ and x^* lies in the interior of Q. Show that x^* is also an optimum solution to $\min\{f(x) \mid x \in P\}$.

[15]

(BQ5) Show that system of inequalities in $\mathbb{R}^n$

$$Ax \leq 0, \quad c^Tx < 0,$$

where $A \in \mathbb{R}^{m\times n}$ and $c \in \mathbb{R}^n$, and the following system of inequalities in $\mathbb{R}^m$

$$A^Ty + c = 0, \quad y \geq 0$$

are strong alternatives. [15]

(BQ6) Let A be a $m \times n$ matrix and $b \in \mathbb{R}^m$ with $b \in \mathcal{R}(A)$.

(a) Let $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$. Show that $c^T x = d$ for all x satisfying $Ax = b$ if and only if there exists $\lambda \in \mathbb{R}^m$ such that $c = A^T\lambda$ and $d = b^T\lambda$.

(b) Show that there exists an x satisfying

$$x \succ 0, \quad Ax = b$$

if and only if there exists no λ with

$$A^T\lambda \succeq 0, \quad A^T\lambda \neq 0, \quad b^T\lambda \leq 0.$$

[9+6=15]

Indian Statistical Institute

End-Semester Examination (2018-2019) Back Paper

M.Tech. (CS) II

Advanced Algorithms for Graph and Combinatorial Optimization Problems

Date: 13.11.2018 Maximum Marks: 100 Time: 3 hours

Answer as much as you can. The maximum you can score is 100 marks. Marks allotted to each question are indicated within square brackets near the right margin. This paper has two pages.

1. Give an example of a chordal graph which is not an interval graph. [5]

2. Write the steps of an algorithm which takes a graph as input, and outputs its perfect elimination order if one exists. [10]

3. Determine a smallest (with respect to the number of vertices) imperfect graph G such that $\chi(G) \neq \omega(G)$, where χ and ω are the chromatic number and the size of the maximum clique of G respectively. [10]

4. (a) Give one example each for a graphic matroid and a matric matroid.
(b) Six employees of a bank visited the vault room on the day when one of the lockers was broken and a rare item was stolen. Each entered once, stayed for some time, and then left. For any two of them that were in the vault at the same time, at least one of them saw the other. Detectives questioned the employees and gathered the following testimony:

STUDENT	CLAIMED TO HAVE SEEN
Anjan	Ranjan, Swapan
Ranjan	Anjan, Tapan
Chandan	Nandan, Tapan
Nandan	Anjan, Tapan
Swapan	Ranjan, Chandan
Tapan	Chandan, Swapan

Assume that the culprit is trying to frame another suspect by providing false information (but not by suppressing information). If one employee did so, who was it? [10+9=19]

5. Prove that a graph G is a split graph if and only if G and $\overline{G}$ are triangulated graphs. [6]

6. Sketch an efficient algorithm, with a brief justification for your claim of efficiency in each case, to (i) recognize whether a given undirected simple graph is a comparability graph, and (ii) report its chromatic number. [10+5=15]

P. T. O

7. (a) Trace the depth-first search based path addition algorithm for planarity testing on the 3-dimensional cube graph Q_3.
   (b) Prove that a maximal planar graph is 3-colorable if and only if it is Eulerian. [7+8=15]

8. Present Edmonds-Karp algorithm to find the max-flow in a given network along with its worst case time complexity. [10]

9. Present a Las Vegas randomized algorithm for comparison based sorting of integers. What is its expected running time? [9+5=14]

10. (a) Define the metric closure of a weighted undirected graph.
    (b) Present an 1.5-factor approximation algorithm for finding a Steiner minimum tree of a given weighted undirected graph $G = (V, E)$ with demand vertices $D \subset V$. Justify the performance guarantee.

[3+(6+7)=16]

INDIAN STATISTICAL INSTITUTE

First Semester Examination : 2018-2019

Course Name : M.TECH. (CS) YEAR II

Subject name : LOGIC FOR COMPUTER SCIENCE

Date 16.11.18: TBA Maximum Marks : 50 Duration : 3 hours

Answer any 10 questions. Notations are used as in the class.

1. Prove that the language of classical propositional logic is countable. [5]

2. Show that the following formulas are theorems of classical propositional logic:

   (a) $\alpha \longrightarrow \neg\neg\alpha$

   (b) $\neg\alpha \longrightarrow (\alpha \longrightarrow \beta)$ [2.5 + 2.5 = 5]

3. Let us consider the binary number system. When we add two numbers of at most two digits, say ab and cd, we get a number of at most three digits, say pqr. Using standard connectives of classical propositional logic, write formulas that express p, q and r in terms of a, b, c and d. [5]

4. Check the satisfiability of the following formulas using the tableau algorithm for classical propositional logic.

   (a) $(((\neg p \vee q) \wedge p) \wedge \neg\neg q)$

   (b) $((p \wedge q) \wedge \neg(p \vee q))$ [3 + 2 = 5]

5. Show that none of the following sentences is logically implied by the other.

   (a) $\forall x \forall y \forall z(Pxy \longrightarrow (Pyz \longrightarrow Pxz))$

   (b) $\forall x \forall y(Pxy \longrightarrow (Pyx \longrightarrow (x = y)))$ [2.5 + 2.5 = 5]

6. Find formulas in prenex normal form equivalent to the following:

   (a) $(\exists x Ax \wedge \exists x Bx) \longrightarrow Cx$

   (b) $\forall x Ax \longleftrightarrow \exists x Bx.$ [2.5 + 2.5 = 5]

7. Assume that a first order language has equality and a two place predicate symbol P. For each of the following conditions, find a sentence φ such that a structure $\mathcal{A}$ is a model of φ iff the following conditions are met. [5]

   (a) The domain of the structure $\mathcal{A}$ has exactly two members.

   (b) The interpretation of P in the structure $\mathcal{A}$, $P^{\mathcal{A}}$ is a function on the domain of $\mathcal{A}$.

8. Consider a first order language $\mathcal{L}$ with equality whose vocabulary consists of only a two-place predicate symbol P. Let $\mathfrak{A}$ be a finite structure corresponding to the language $\mathcal{L}$. Let $\mathfrak{B}$ be another structure corresponding to the language $\mathcal{L}$ such that $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent. Are these two structures isomorphic ? Justify your answer. [5]

9. There exists a set of sentences in first order logic whose models are precisely the finite sets. Prove or disprove. [5]

10. Let $\mathcal{L}$ denote a countable first order language and T denote a theory in $\mathcal{L}$. Show that if T is axiomatizable and complete, then T is decidable. [5]

11. What are the definable subsets of $\mathbb{N}$ in $\mathfrak{N}_L = (\mathbb{N}, 0, S, <)$? Give justifications. [5]

12. Show that the relation $\mathcal{M} = \{(m, n, p) : m, n, p \in \mathbb{N}, \text{ and } p = m \times n\}$ is not definable in Presburger Arithmetic. [5]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2018-2019

M. Tech. (CS) II year

Data Mining and Knowledge Discovery

Date: 16 .11.2018 **Maximum Marks: 100** **Duration: 3 hours**

[Answer as much as you can]

1. (i) Enumerate three major clustering approaches, indicating a few examples of each.
   (ii) How do we define similarity in the context of hierarchical clustering?
   (iii) How do we validate a clustering result? Define some such indices.

   [6+5+6=17]

2. (i) Comment on the termination criteria for hierarchical clustering.
   (ii) Describe algorithm DBSCAN.
   (iii) What are some of the limitations of partitive algorithms like k-means, and how can one overcome these?

   [4+8+7=19]

3. (i) How does FP-tree help in rule mining? Explain tree construction with an example.
   (ii) Describe algorithm PAM, and analyse its swapping process.
   (iii) What is overfitting? How is it different from underfitting?
   (iv) Outline some stopping criteria for a classifier. [9+6+5+4=24]

4. (i) Given a trained logistic classifier for a single feature and 2 classes, what is the equation for the decision boundary if W=2 and b=1?

(ii) How can we generalize logistic classification to more than 2 classes?

(iii) What are the problems of using a sigmoid activation function?

(iv) Discuss briefly why scaling of the input features is important.

(v) Why is initializing all the weights to zero problematic?

(vi) Explain briefly how momentum gradient descent works, and why this can be more robust than regular gradient descent.

(vii) Given a 32x32x5 image, you filter it with a 5x5x5 kernel in the way most Convolutional Neural Networks are implemented. If you use no padding, what will be the output size of the activation map?

[2+2+2+2+2+4+1=15]

P. T. O.

5. (i) What does it mean to use a "Fully-convolutional" architecture for image segmentation?

   (ii) How does batch size relate to learning rate? Explain

   (iii) Why is it a problem to optimize accuracy directly with a deep neural network?

   (iv) What is the reasoning behind the concatenation operations in U-Net for image segmentation?

   (v) You have a convolutional neural network trained for image classification. Describe a simple way of detecting what parts of an image are responsible for a certain classification result, without using the image gradients.

   (vi) Give two possible explanations as to why residual networks work better than standard feed forward networks.

   [3+3+3+3+5+4=21]

6. Write short notes on the following:

   (i) Capsule network

   (ii) Generative Adversarial Network

   (iii) Long Short Term Memory

   (iv) VC Dimension [5+5+5+5=20]

Indian Statistical Institute

Semester-1 2018-2019

M.Tech.(CS) - Second Year

End-semester Examination (November 19, 2018)

Subject: Computer Architecture

Maximum marks: 100 Duration 3 hrs.

Please keep your answers brief and to the point.

Answer any 4 questions. Each question carries 25 marks.

1. (a) Explain the purpose of the *activate* and *precharge* commands in the context of DRAMs. [4]

   (b) Explain the advantage of using a virtually-indexed physically-tagged (VIPT) L1 cache. [2]

   (c) Consider a demand paging system with a physical memory of 4GB and a virtual page size of 4KB. For the page replacement policy, we wish to evaluate the *true LRU* and the *Victim-Next-Victim* strategies.

   (i) Find the minimum number of bits required to store the necessary information to implement the true LRU policy. Where should these bits be stored? [3+2]

   (ii) Find the minimum number of bits required to store the necessary information to enable the *Victim-Next-Victim* policy. [3]

   (iii) Can you conclude which among *LRU* and *Victim-Next-Victim* works better in this case? Explain your answer. [5]

   (d) Explain why modern DRAMS are organized into banks. Explain the basic difference between the banked structures of the $L1$ cache and the DRAM. [4 + 2]

2. (a) A processor implements an inorder pipeline with 12 stages. Each stage completes in a single clock cycle. The pipeline stalls on a conditional branch instruction until the condition of the branch is evaluated. However, you *do not know* at which stage the branch condition is evaluated. The processor is known not to employ any prediction mechanism. A program with 1000 dynamic instructions completes in 2211 cycles. If 200 of those instructions are conditional branches, at the end of which pipeline stage the branch instructions are resolved? Assume the pipeline does not stall for any other reason than the conditional branches during the execution of that program. [6]

   (b) Explain the working principle of Tomasulo's algorithm with an example. [8]

   (c) Explain how the non-speculative Tomasulo algorithm resolves the following classes of hazards: **RAW, RAR, WAW, WAR.** [2+2+2]

   (d) Explain the differences between dependencies through registers and dependencies through memory. [5]

3. (a) Comment on the following: (i) Small cache blocks versus Sectored caches (ii) Virtually addressed cache versus physically addressed cache (iii) Large versus small pages [3+3+3]

   (b) Can a cache be 5-way set associative? Justify your answer. [2]

(c) A processor has a 4-way set associative L1 cache that can accommodate 4 blocks in total. The access latency to this cache is 1 cycle. The replacement policy is true LRU. The processor is known not to employ any prefetching mechanism. The processor also has a 16-way set associative L2 cache that can accommodate 128 blocks in total. The access latency to this cache is 20 cycles. A programmer writes a test program that in a loop repeatedly accesses only the following data cache blocks (assume billions of iterations are run):

$$A, B, C, D, E, F$$

where A, B, C, D, E, F are different cache block addresses. In the steady state (i.e. after the loop has executed for a few iterations), the programmer finds that the average memory access time is 1 cycle. Then, the programmer writes another program that in a loop repeatedly accesses only the following data cache blocks:

$$A, B, C, D, E, F, G, H$$

In the steady state (i.e. after the loop has executed for a few iterations), the programmer finds out that the average memory access time is 20 cycles.

(i) Can you explain the disparity in average memory access times between the above two cases? In other words, can you comment on the cache organization of the processor? [3]

(ii) Based on the above information, what do you expect the average memory access time to be for another program that in a loop repeatedly accesses only the following data cache blocks? [2]

$$A, B, C, D, E$$

(iii) Again, based on the above information, comment on the average memory access time for yet another program that in a loop repeatedly accesses only the following data cache blocks: [2]

$$A, B, C, D, E, F, G$$

(d) Consider the Belady's OPT cache replacement algorithm, which replaces the victim block such that it is going to be referenced furthest in the future by the program.

(i) How would you design a processor that implements this policy? [4]

(ii) Is this algorithm optimal in minimizing cache miss rate? Explain your answer. [3]

4. (a) Consider a multiprocessor system with private $L1$ caches and a shared $L2$ cache, in which shared data is read by all cores every cycle, but written once per 1000 cycles by a single core. Comment which among $UPDATE$ or $INVALIDATE$ would be a better choice for a coherence protocol for such a system. Justify your choice. [4]

(b) Assume we use the directory based cache coherence mechanism. Consider at a particular instant, for a particular cache block A, the bit vector stored in the directory contains all zeroes. What does this tell you about the cache block? [2]

(c) Consider a multi-processor system with 512 processors. Each processor has a 1 MB private writeback cache with 64-byte cache blocks. The main memory size is 1 GB.

(i) If we implement a snoopy bus based $MESI$ cache coherence protocol, how many bits of state do we need in the entire system for the coherence protocol implementation? Where do these bits reside? [4+2]

(ii) Consider we decide to implement a *hybrid directory* cache coherence protocol with each processor keeping $MESI$ bits (instead of a single exclusive bit per block). How many bits do we need in the entire system for the coherence protocol implementation? Where do these bits reside? [4+2]

(iii) Which of the above protocols would you choose for this system? Justify your answer. [3]

(d) Consider a loosely coupled multiprocessor system. What type of cache coherence protocol you would choose for such a system? Justify your answer. [4]

5. (a) Consider the design of an SIMD engine that can support a vector of length 16. We have two options: (i) A traditional vector processor (ii) A traditional array processor.
Which one is more costly in terms of chip area? Justify. [3]

(b) Consider a program running on a vector processor with the following latencies for various instructions:

- VLD and VST: 50 cycles for each vector element; fully interleaved and pipelined
- VADD: 4 cycles for each vector element (fully pipelined)
- VMUL: 16 cycles for each vector element (fully pipelined)
- VDIV: 32 cycles for each vector element (fully pipelined)
- VRSHF (right shift): 1 cycle for each vector element (fully pipelined)

Assume that:

- The machine has an in-order pipeline
- The machine supports chaining between vector functional units
- In order to support 1-cycle memory access after the first element in a vector, the machine interleaves vector elements across memory banks. All vectors are stored in memory with the first element mapped to bank 0, the second element mapped to bank 1, and so on.
- Each memory bank has an 8KB row buffer
- Vector elements are 64 bits in size
- Each memory bank has two ports (so that loads / stores can be active simultaneously), and there are two load / store functional units available.

(i) What is the minimum power-of-two number of banks required in order for memory accesses to never stall? Assume a vector stride of 1.

(ii) The machine (with as many banks as you found above) executes the following program (assume again vector stride set to 1):

```
VLD V1 <- A
VLD V2 <- B
VADD V3 <- V1, V2
VMUL V4 <- V3, V1
VRSHF V5 <- V4, 2
```

It takes 111 cycles to execute this program. What is the vector length?

(iii) If the machine did not support chaining (but could still pipeline independent operations), how many cycles would be required to execute the same program? Justify your answer. [4+6+6]

(c) Explain briefly on the example above or otherwise, the performance benefits achieved by a vector processor with (i) Banked memory (ii) Multiported memory. [3+3]

6. (a) Assume that we test the performance of two processors, A and B, on a benchmark program. We find the following about each:

- Processor A has a CPI of 2 and executes 4 Billion instructions per second
- Processor B has a CPI of 1 and executes 8 Billion instructions per second

Which of the above processors has a better performance on this benchmark? Recall that CPI stands for average number of Cycles Per Instruction. [3]

(b) Value prediction is used as a method to handle data dependences. One method of value prediction for an instruction is *last-time prediction*. The idea is to predict the value to be produced by the instruction as the value produced by the same instruction the last time the instruction was executed. If the instruction was never executed before, the predictor predicts the value to be 1. Value prediction accuracy of an instruction refers to the fraction of times the value of an instruction is correctly predicted out of all times the instruction is executed.

Consider the following fragment of code, which has 4 load instructions in each loop iteration, loads to arrays x, y, z, and t.

```
// Initialize integer variables c, d, e, f to 0
// Initialize integer arrays x, y, z, t
for (i = 0; i < 1000; i++) {
    c += x[i];
    d += y[i];
    e += z[i];
    f += t[i];
}
```

Assume the following state of arrays before the loop starts executing:

- x consists of all 0's.
- y consists of alternating 3's and 6's in consecutive elements.
- z consists of random values between 0 and 2^{32} - 1.
- t consists of 0,1,2,3,4,....,999.

(i) What is the value prediction accuracy of the aforementioned predictor for the four load instructions in the program? [2+2+2+2]

A. Load x[i]

B. Load y[i]

C. Load z[i]

D. Load t[i]

(ii) Justify if you can design a predictor that can achieve higher accuracy for the prediction of x[i]. [2]

(iii) Justify if you can design a predictor that can achieve higher accuracy for the prediction of y[i]. [2]

(iv) Justify if you can design a predictor that can achieve higher accuracy for the prediction of z[i]. [2]

(v) Justify if you can design a predictor that can achieve higher accuracy for the prediction of t[i]. [2]

(c) You are a programmer in a large corporation and you have been asked to parallelize an old program so that it runs faster on modern multicore processors.

(i) You parallelize the program and discover that its speedup over the single-threaded version of the same program is significantly less than the number of processors. You find that many cache invalidations are occurring in each core's data cache. What program behaviour could be causing these invalidations?

(ii) You modify the program to fix the performance issue. However, now you find that the program is slowed down by a global state update that must happen in only a single thread after every parallel computation. In particular, your program performs 90% of its work (measured as processor-seconds) in the parallel portion and 10% of its work in this serial portion. The parallel portion is perfectly parallelizable. What is the maximum speedup of the program if the multicore processor had an infinite number of cores? How many processors are required to attain a speedup of 4? [1+3+2]

M.Tech (CS) -2nd Year

Semester Exam - 2018

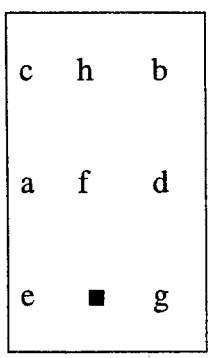
Artificial Intelligence

19.11.2018

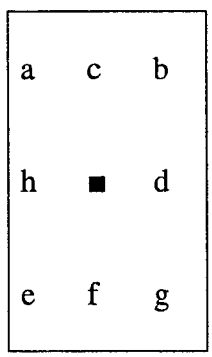
Full Marks - 100 Duration - 3 hours

Answer Any Five.

1. Let us consider the following 8-puzzle problem. The start node and goal node configurations are given below:



start node



goal node

Consider a simple evaluation function

$$f(n_i) = d(n_i) + w(n_i)$$

where $d(n_i)$ is the depth of node n_i in the search tree and $w(n_i)$ is the number of misplaced tiles in the node n_i with respect to the tiles of the goal node. Find out the solution of the 8-puzzle problem using A*- search algorithm.

(20)

2. Prove that the following set 'S' of clauses is unsatisfiable.

$$S \equiv \{\sim P \vee \sim Q \vee \sim R, \sim P \vee S, P \vee S, Q \vee R, \sim Q \vee S, \sim Q \vee P, P, Q, \sim R, \sim P \vee \sim Q \vee R\};$$

where, the set of atoms is {P, Q, R, S}.

(20)

3. Consider the following two clauses:

$C_1 = P(x, h(a, y)) \vee Q(f(a), y) \vee \sim R(x, z, b) \vee \sim R(x, h(x,x), b);$

P T O.

$C_2 = \sim P(f(a), h(u,v)) \vee \sim Q(v, g(u)) \vee R(u, u, b)$

Find all possible binary resolvents of the above two clauses. (20)

4. a. What is the difference between McDermott's Nonmonotonic Logic and Moore's AutoEpistemic Logic from semantical point of view?

b. Consider the sentence - "Peter usually comes late at work if there is not any important meeting". Express this sentence by means of a default rule.

c. Consider the following set of rules in McDermott and Doyle's NML:

$$A = \{\mathbf{M}p \supset \sim r, \mathbf{M}r \supset \sim p\}.$$

Determine the fixed point(s).

d. State the properties of *extension* of a default theory T = (W,D).

e. Consider the default theory $\boldsymbol{T = (W, D)}$ with $\boldsymbol{W}$ = {green, aaamember} and $\boldsymbol{D}$ containing the following defaults:

$$\delta_1 = \frac{green:\neg \text{likesCars}}{\neg \text{likesCars}}, \quad \delta_2 = \frac{aaaMember:\text{likesCars}}{\text{likesCars}}$$

Using process tree, derive the extension(s) of $\boldsymbol{T}$.

(2+2+5+3+8 = 20)

5. a. Consider the following program:

$$fly(X) \leftarrow bird(X), \mathbf{not}\ \neg fly(X).$$
$$\neg fly(X) \leftarrow penguin(X).$$
$$bird(tweety) \leftarrow.$$
$$bird(rocky) \leftarrow.$$
$$penguin(rocky) \leftarrow.$$

Write the grounded version of this program. Verify whether the following set is an answer set of the program

$$\{bird(tweety), bird(rocky), penguin(rocky), fly(tweety), \neg fly(rocky)\}.$$

b. Derive the Answer set of the following program:

P.T.O

$p \leftarrow a.$
$a \leftarrow \textbf{not } b.$
$b \leftarrow \textbf{not } a.$

(12 + 8 = 20)

6. a. What are the advantages of Mini-Batch Gradient Descent over Stochastic Gradient Descent and Batch Gradient Descent?

b. What are the different ways the hyper-parameters of a neural network can be tuned? Explain briefly.

c. Which features do the first layers learn in a CNN? Why are they said to be *general*? What are the reasons for drop in performance in Transfer Learning? Does fine-tuning transferred layers perform better than frozen transferred layer? Why?

d. What are the various techniques to handle Imbalanced datasets? Name them.

(5+4+(1+1+3+3)+3 = 20)

7. a. Draw a Tensor Graph for a 2-layer CNN for binary classification. Clearly show convolution, pooling, activation functions and other required operations.

b. Differentiate between Max Pooling and Average Pooling.

c. What is One-shot learning?

d. What are the advantages of Bi-RNN over RNN?

e. Can Image Captioning be done using LSTM? If Yes, show a flow diagram and explain the working of the different blocks. If No, explain why.

(6+4+3+3+4 = 20)

INDIAN STATISTICAL INSTITUTE

Semestral Examination : (2018 - 2019)

Course Name : M. Tech. (CS) **Year : 2nd year**

Subject Name : Neural Networks & Applications

Date : November 22, 2018 **Maximum Marks : 100** **Duration : 3 hrs 30 mins**

Answer questions {1, 3, 4} OR {2, 3, 4}.

1. Derive the expressions for computing the amount of modification for the weight values of the links, in each iteration during training under backpropagation learning algorithm, in a multilayer perceptron for the purpose of pattern classification. Assume two hidden layers between the input and output layers, and also consider appropriate non-linear activation functions of the computing nodes, and the energy function at the output layer. [20]

2. Consider a two-input XOR problem in the context of pattern classification.
   (a) Show that a perceptron model with a single linear node, where an identity function is used as its transfer function, cannot perform the task of classification for this problem. [8]

   (b) The above model is then modified by introducing two nodes in a hidden layer with ReLU as their activation functions to build a multilayer perceptron model. Note that there is only one output node as in the above perceptron model. Show how this model is now able to perform the task of pattern classification. [12]

3. (a) Consider a set of 10×10 RGB images distributed in 2 classes. You have to classify these images using either a multilayer perceptron or a convolutional neural network with appropriate architecture. Explain with reasons with appropriate numerical examples, which one you would select in this regard so that computational burden becomes minimum. [10]

(b) State and explain different hyper-parameters used in a convolutional neural network. Write down the relations among these hyper-parameters with appropriate explanation. [10]

(c) Consider a set of $n \times n$ multichannel images distributed in several classes. These images need to be classified using a convolutional neural network with two convolution layers and a fully connected layer, apart from input and output layers. The first convolution layer corresponds to filtered images of size $n \times n$ each. The second convolution layer downsizes these filtered images to ones with size $n/2 \times n/2$ each. State and explain the expression for the amount of modification, in each iteration during training under gradient-descent minimization, for the weight values of the links between the input and the first convolution layers. [20]

4. Consider a time series prediction problem, where you need to predict l-dimensional output from n-dimensional time series data using a simple recurrent neural network (RNN). Assume that the maximum allowable time is 10 during which the outputs are fed back appropriately under the framework of the RNN. That is, the network is able to model the dynamical system generating this time series data, which is defined as

$$\mathbf{s}(t) = \mathbf{f}(\mathbf{s}(t\text{-}1), \mathbf{x}(t); \boldsymbol{\theta})$$

Here, $\mathbf{s}(t)$ is the state of the system at time t, $\mathbf{x}(t)$ is the input signal, and $\boldsymbol{\theta}$ is the parameter of the model.

(a) Derive an expression for the gradient of the error function of the RNN with respect to the input to the hidden nodes corresponding to 5^{th} unit time, under backpropagation through time (BPTT). [15]

(b) Derive the learning rules for modification of all the weights values of the links present in the RNN under BPTT. [20]

(c) Hence explain the problems of vanishing and exploding gradient that may arise in such a neural network model. [5]

INDIAN STATISTICAL INSTITUTE

Semestral Examination:(2018-2019)

M.TECH (CS) II YEAR

Subject Name: Quantum Information Processing and Quantum Computation

Maximum Marks: 60 Duration:3 hours Date: 24·11·2018

Answer any five of the following questions

1. a) Consider the following two-qubit state shared between Alice, Bob at distant laboratories;

$$|\psi\rangle_{AB}^{-} = \frac{1}{\sqrt{2}}[|01\rangle - |10\rangle]$$

$|0\rangle$ and $|1\rangle$ form an orthonornal basis in two dimensional Hilbert space.
i) Show that a known state $a|0\rangle + b|1\rangle$ with a, b real, can be prepared in a remote lab of Bob if Alice and Bob shared the above state.
ii) Prove that the if both the parties measure the spin in the same direction, there result is perfectly anti-correlated.
iii) Show that this quantum state violates Bell's inequality for suitable choice of measurements.

[2+4+6]

2. i) Consider a two qubits pure state $|\psi\rangle_{12}$ and four unitary operators $\{U_i, i = 1, 2, 3, 4\}$ acting on the first particle. What is the necessary condition so that the four states $\{U_i \otimes I|\psi\rangle_{12}, i = 1, 2, 3, 4\}$ form an orthogonal set?
ii) Find a two qubits state $|\psi\rangle_{12}$ for which $\{U_i \otimes I|\psi\rangle_{12}, i = 1, 2, 3, 4\}$ form an orthogonal set.
iii) Discuss how quantum super dense coding can be realized using the state $|\psi\rangle_{12}$ having the properties described in (ii).

[3 + 3 + 6]

3. a) Let a single qubit unitary gate U be realized in the following way;

$$U = A\sigma_x B\sigma_x C$$

P.T.O

where A, B and C are also single qubit gates with $ABC = I$ Show using diagram, how two-qubit gate $C - U$ can be implemented by single qubit gates and $C - NOT$ gates.
b) Let V and U are two gates where $U = V^2$. Draw the circuit diagram and analyze how the two-qubit $C^2 - U$ can be implemented by elementary gates.
c) Consider a function $f_a : \{0,1\}^n \rightarrow \{0,1\}$ where the function f_a is given by $f_a(x) = x.a$, a being a n-bit string. On how many points the function has to be calculated to determine a? Describe the faster quantum algorithm by which a can be determined.

[3 + 3 + 6]

4. Consider a function $f : \{0,1\}^n \rightarrow \{0,1\}^n$. The function has a period given by n-bit string a: that is
$f(x) = f(y)$ iff $y = x \oplus a$
a) Discuss how hard it is to find the period a in the classical world.
(b) Discuss the quantum algorithm by which the period can be found in polynomial time.

[2+10]

5. The function $f : \{0,1\}^n \rightarrow \{0,1\}$ is such that;
$f(x) = 1$, for $x = \omega$ and $f(x) = 0$ for $x \neq \omega$
a) In the classical world, how many queries are required to find ω?
b) Show how a quantum algorithm can provide a quadratic speed up for this search problem.

[2+10]

6. a) Discuss how phase error in computational basis

$$|0\rangle \rightarrow |0\rangle \quad and \quad |1\rangle \rightarrow -|1\rangle$$

can be corrected.

b) Consider the following two nine-qubit orthogonal states

$$|0\rangle_L = \frac{1}{2\sqrt{2}}[|000\rangle + |111\rangle][|000\rangle + |111\rangle][|000\rangle + |111\rangle]$$

$$|1\rangle_L = \frac{1}{2\sqrt{2}}[|000\rangle - |111\rangle][|000\rangle - |111\rangle][|000\rangle - |111\rangle]$$

Show that using this code, any one qubit error can be corrected.

[4 + 8]

INDIAN STATISTICAL INSTITUTE

End-Semestral Examination: 2018

Date : 26.11.2018
Subject Name : **Cryptology**
Course Name : M.Tech. (CS) II yr. Max Score: 60 Duration: 180 Mins
Note: Attempt all questions. Marks are given in brackets. Total score is 70. But maximum you can score is 60. Use separate page for each question.

1.[3 + 5 = 8] Define the computational distance, denoted $\Delta_{\mathcal{A}}(\mathsf{X};\mathsf{Y})$, between two random variables X and Y over a finite set Ω, with respect to a PPT algorithm $\mathcal{A}$. Give an example of Ω, X and Y, such that $\Delta_{\mathcal{A}}(\mathsf{X};\mathsf{Y}) \approx 0$ for all PPT algorithm $\mathcal{A}$, and $\Delta(\mathsf{X};\mathsf{Y}) \approx 1$, where $\Delta(\mathsf{X};\mathsf{Y})$ denotes the statistical distance between X and Y. Clearly state all the hardness assumptions, if any.

2.[3 + 5 = 8] Define the CTR mode of operation based on a keyed function $F_K : \{0,1\}^n \to \{0,1\}^n$. For an n-bit block cipher E_K, we define $F_K(x) = E_K(x) \oplus x$, for all $x \in \{0,1\}^n$. Construct an $O(2^{n/2})$-query nonce-respecting distinguisher for CTR mode based on such F_K.

3.[3 + 7 = 10] For $r \geq 1$, define the r-round Feistel network based on keyed functions $F_{K_i} : \{0,1\}^n \to \{0,1\}^n$, where $1 \leq i \leq r$. Construct an $O(1)$-query SPRP distinguisher on 3-round Feistel network when F_{K_i} are independently keyed block ciphers.

4.[2 + 4 + 6 = 12] Define ϵ-universal hash functions. Show that $\mathsf{polyHash}_K : (\{0,1\}^n)^{\leq L} \to \{0,1\}^n$ is an $L/2^n$-universal hash function, when $K \leftarrow_{\$} \{0,1\}^n$. Construct an $O(2^{n/2})$-query forgery attack on $E_{K'} \circ \mathsf{polyHash}_K$, where $E_{K'}$ is an independently keyed n-bit block cipher.

5. [5 + 5 = 10] Let E_K be an n-bit blockcipher with n bit key.

1. Construct a $O(1)$ complexity collision finding algorithm for $f_1(h, m) = E_{h \oplus m}(m) \oplus h \oplus m$.
2. Construct a $O(2^{n/2})$ complexity preimage finding algorithm for the compression function $f_2 : \{0,1\}^{3n} \to \{0,1\}^n$ where $f_2(h, h', m) = E_h(E_{h'}(m)) \oplus m$.

6 [2+5+3=10] Let p be a prime such that $(p-1)/2$ is also a prime. Let g be a generator of the group Z_n^*. Given any fixed h, we define a function $f(a, b) = g^a h^b$ where $0 < a, b < p$. Show that it is indeed a compression function. If A is an algorithm which can find a random collision of f with probability one, then show that there is an efficient discrete log problem solver of Z_p^*. Compute the success probability of your algorithm solving discrete log problem.

7 [4+ 4 +2 + 2 = 12] State Diffie-Hellman Key Exchange. Construct an active adversary (which can modify the communication between users) of the key exchange. What are the secrets of the users in the presence of the active adversary. Show that the secret keys are known to the adversary.

Indian Statistical Institute
End-Semester Examination (2018)
M. Tech.(CS) II Year
Cognitive Science

Date: 26.11.2018 **Full Marks: 80** **Duration – 3 hours**

Answer as many questions as you like, but you can at most score 80

1. (a) "Until recent past, neuroscientists had to wait for disaster to strike people and if the victims pulled through, see how their minds worked differently afterward." Do you agree with this statement? Justify your answer in the light of the experiences of the two legendary neurosurgeons Ambroise Pare & Andreas Vesalius.

(b) Hence explain some limitations of the 1400 years old theory of Galen of Pergamon that Andreas Vesalius pointed out.

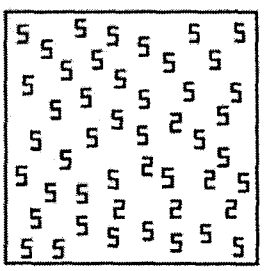
(c) State the contributions of Camillo Golgi and Ramon y Cajal that helped neuroscience to advance beyond the limitations mentioned in (a) above. What was the precondition to such contribution? (1+5)+4+5=15

2. (a) What is the difference between perception and cognition?

(b) Do you consider Synesthaesia a cognitive disability? Explain your answer.

(c) In visual search, explain the difference between Serial Search & Pop-out Search. If you are given the task of plotting *number of distracters* vs *search time* for serial search and pop-out search respectively, draw the shape of the two curves with justification.

(d) Explain in the light of **(c)** above, what happens in case of the following picture for someone with Grapheme-colour Synesthaesia.



3+(1+4)+(4+3)+3=18

3. (a) What is Neglect disorder in cognition?

(b) In the light of this disorder explain the following statement "After suffering a stroke in 1967, the German painter Anton Radersheidt had to relearn the act of painting".

(c) What truth regarding visual space as well as memory was unveiled from the experiments of the well-known Indian neuroscientist V S Ramachandran regarding the neglect disorder leading to Anosognosia?

(d) What truth regarding the human brain anatomy and faulty communication in signaling pathway therein is revealed from Sleep Paralysis? 3+4+5+6=18

PTO

4. (a) Do you agree with the following statement: "Antagonistic receptive fields are the key to human perception of sensory experiences." Justify your answer.

(b) What is the most popular mathematical model of the antagonistic receptive fields?

(c) What signal do the mathematical functions used in the model physically represent?

(d) Hence briefly outline the ionic basis of generation of such a signal in a neuron using the equation for Nernst potential and drawing the corresponding equivalent electrical circuit/s.

(e) How is the physiological phenomenon in (d) modeled in Artificial/Machine Intelligence domain? (1+4)+3+2+8+4=22

5. (a) Define Intelligence as suggested by Behaviourists/Cognitive Scientists.

(b) How does renowned computer scientist Marvin Minsky algorithmically explain child intelligence using Building Block model?

(c) Can the explanation in (b) extended to the working of any complex network as in ant colony functioning and as in connecting the brain neurons? Justify your answer.

(d) Explain the relevance of algorithms in (c) in the above context, considering C M Bhatia's Pattern Drawing Test in Cognitive Science on one hand, and the Eulerian path printing in Graph theory, on the other. 3+5+(1+5)+6=20

6. **(a)** In his book *The Descent of Man, and Selection in Relation to Sex*, explain Charles Darwin's viewpoint on development of human cognitive skills with respect to use of voice like, for example, singing?

(b) Explain the *Baldwin Effect* in the course of evolution of cognitive skills. Does it support Lamarckism as opposed to Darwin's theory? Justify. 4+(4+1+3)=12

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2018-19

M. Tech. (Computer Science) Second Year

Natural Language Processing

Date: 28. 11. 2018 **Maximum Marks: 50** **Duration: 2 h 30 min**

NOTE: **Total Marks: 60,** Answer all questions.

Write answers on question paper itself. You may use the usual answer script for rough work.

Roll No. _____

---

1. (a) Why is Multi-Layer Perceptron (MLP) not suitable for NLP tasks? **[5]**

(b) How does a recurrent architecture help in solving the problem faced by MLP? **[5]**

2. (a) Why is **linear softmax** not suitable for skip-gram and continuous bag of words (CBOW) models? **[5]**

(b) How do **word2vec** architectures ensure that similar words get similar vectors? **[5]**

3. (a) Why is LSTM better than RNN in capturing long term dependencies? **[5]**

(b) If the cell update equation of LSTM is changed from $C_t = C_{t-1} + i_t \times \tilde{C}_t$ to $C_t = C_{t-1} \times i_t \times \tilde{C}_t$, will the new LSTM suffer from vanishing gradient? Justify. **[5]**

4. (a) We write dates in many different formats which are to be normalized to a fixed format. Example:

May 12th 1980	=>	12-05-1980	
21 Feb., 1959	=>	21-02-1959	
15th August, 1947	=>	15-08-1947	etc.

Design an RNN based system that takes dates in different formats and normalizes them. Give diagram and write equations for your system. Briefly explain the functioning of the system using an example. **[15]**

(b) Does the number of parameters of your system change if only *day* and *month* are given in the input? Justify your answer. **[5]**

5. Assume that a sufficiently large language corpus is divided into six parts (r_1, r_2, ..., r_6). You observe the numbers of pronouns in three particular parts (say, in r_i, r_j and r_k) are 200, 100 and 300. The observed mean($\overline{n}$) is 200. For distribution of pronouns, let p_i denote the probability that a pronoun belongs to r_i and n_i denote the number of pronouns observed in r_i. Find one possible distribution (write down the values of p_i's and n_i's) of the pronouns in the six parts which is consistent with the following two constraints and the distribution should have **maximum entropy** over the other possible distributions:

$$\text{Constraint 1: } \overline{n} = \sum_{i=1}^{6} n_i p_i \text{ ; Constraint 2: } \sum_{i=1}^{6} p_i = 1$$

Justify your answer. **[10]**

INDIAN STATISTICAL INSTITUTE

Third Semester Examination 2018-19

M.Tech.(CS) - Second Year

Subject: Pattern Recognition and Image Processing

Date: 30.11.18 Total Marks: 110 Maximum Marks: 100 Duration: 3 hours.

Please keep your answers brief and to the point.
Question 1 carries 10 marks and Questions 2-6 carry 20 marks each
Maximum you can score is 100

1.

a. Assume that there are c classes $w_1, w_2, \ldots, w_c$, and an input pattern vector x. Give the Bayes' rule for classification in terms of a priori probabilities of the classes and class conditional probability densities of x. Explain all symbols and their significance. [5]

b. Suppose we have two-classes (A, ~ A), with a single binary valued feature x taking values {0, 1}. Assume the prior probability P(A) = 0.33. Given the distribution of the samples as shown in the following table, use Bayes' rule to compute the values of posterior probabilities of the classes.

[5]

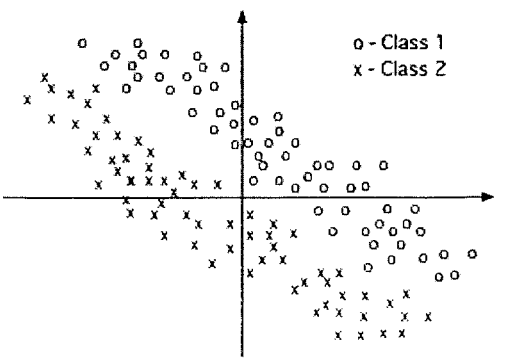
	A	~A
0	248	167
1	82	503

2.

a. Use single-linkage, complete-linkage and average-linkage agglomerative clustering algorithms to cluster the following 8 examples: [10]

A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).

b. Suppose, you are given 2D feature vectors for a classification task which are distributed according to the figure below.



PTO

i. On the figure, draw the location of the mean and the direction of the eigenvector which corresponds to the largest eigen value (no need to show the mathematics) for each class. [2]

ii. Explain briefly how you might use the results of PCA on each class separately to perform classification. [2]

iii. Specify the steps of the PCA algorithm. [6]

3.

a. What is log transformation? How is it useful in image processing? [4]

b. What is image compression? Why is it needed? Explain. [4]

c. Explain the effect of noise on edge detection. [4]

d. Write an expression for 2-D Discrete Fourier Transform. What is its relation with 1D- DFT? [8]

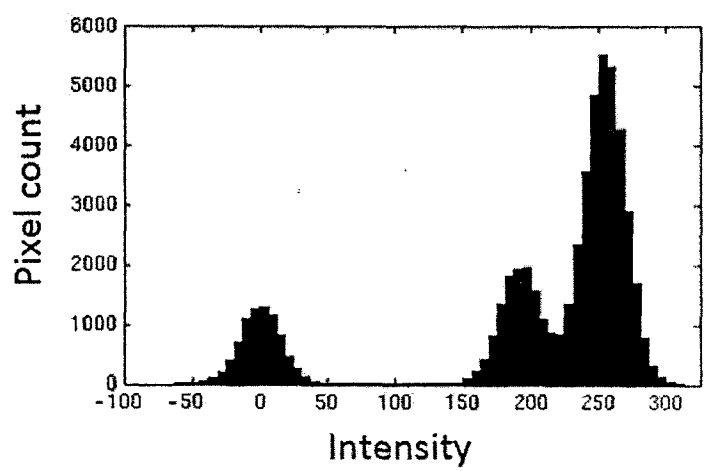
4.

a. Define neighborhood of a pixel. [2]

b. What is meant by image segmentation? Describe its use in image processing. [4]

c. What is the principle of region growing based image segmentation? [4]

d. From the histogram of an image given below, comment on how many regions you can get after segmenting the image. [2]



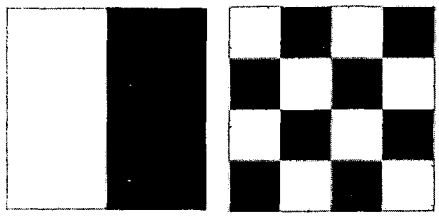
e. An image is to be segmented by a thresholding technique.

i. Describe a process you could use to determine a good value for the threshold. [4]

ii. Describe an automated process that could determine a good value for the threshold. [4]

5.

a. The (binary) images shown below are quite different, but their histograms are the same. Suppose that each image is blurred with a 3 × 3 averaging kernel.



i. Would the histograms of the blurred images still be equal? Explain. [3]

ii. Sketch the two new histograms. [3]

b. Given a small image of 16 × 16 pixel, 3-bit grayscales, with frequency:

count(0) = 0
count(1) = 2
count(2) = 1
count(3) = 69
count(4) = 5
count(5) = 179
count(6) = 0
count(7) = 0

where count(x) = n means that there are n pixels with intensity x in the image.

i. Stretch the histogram using histogram equalization. [8]

ii. Is this transformation reversible? Explain. [2]

c. Explain how contrast stretching is performed. [4]

6.

a. There are three image masks with size 3x3, named W_a, W_b and W_c, given as

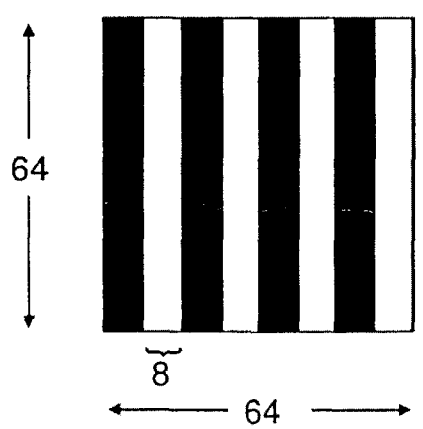
$$W_a = \frac{1}{6}\begin{bmatrix}0 & 1 & 0\\ 1 & 2 & 1\\ 0 & 1 & 0\end{bmatrix}, W_b = \begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix}, W_c = \begin{bmatrix}1 & 1 & 1\\ 1 & -8 & 1\\ 1 & 1 & 1\end{bmatrix}$$

Characterize the three masks by looking at their coefficients. Also describe what the individual masks will do to an image f and the applications they normally have. [6]

b. Given a 64x64 image of black (0) and white (255) vertical stripes (each strip is of width 8) :

The image is convolved with:

$$\frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix}.$$



How will the image look like after the operation? [14]

Indian Statistical Institute
Semestral Examination: 2018
Course Name: M. Tech. in Computer Science
Subject Name: Mobile Computing

Date: 01.12.18 Maximum Marks: 100 Duration: 3 hours 30 minutes

Instructions: You **may** attempt **all** questions which carry a total of **110** marks. The maximum marks you can score is only **100**.

1. (a) How does fixed spectrum access policy contribute to the spectrum scarcity problem in cognitive radio networks? [5]

   (b) What are the different types of cognitive capabilities with which a cognitive radio user should be equipped to support dynamic spectrum access? [5]

   (c) Describe briefly the opportunistic and concurrent spectrum access models in cognitive radio networks. [5+5=10]

   (d) How *matched filter detection* technique is used for indirect spectrum sensing in cognitive radio networks? [5]

   (e) Describe a direct spectrum sensing technique in cognitive radio networks. [5]

2. (a) What is attribute-based addressing in wireless sensor networks? [5]

   (b) What are the differences between the *self-diagnosis* and *cooperative diagnosis* fault detection techniques in wireless sensor networks? [5]

   (c) Describe how clusters are formed and energy-usage among the nodes are balanced in low-energy adaptive clustering hierarchy (LEACH) routing protocol in wireless sensor networks. [10]

   (d) Consider a linear network consisting of n sensor nodes and the base station (BS) as shown in Figure 1 where the distance between two consecutive nodes is r. The distance between the last node n and the BS is also r. Circles denote the sensor nodes and the square denotes the BS.

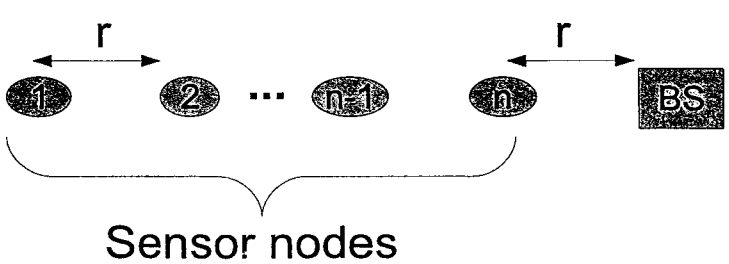


Figure 1: Linear network with n nodes and the base station.

Derive expressions for total energy expended in the system for transmitting a k-bit message from node 1 to the BS using direct communication and minimum transmission energy (MTE) routing protocols. [5+5=10]

(e) Consider the following minimum relay node placement problem in wireless sensor networks: Given a set of sensor nodes S with their locations and an uniform communication

P. T. O.

radius d of both sensor and relay nodes, the problem is to place a set of relay nodes R such that the whole network G consisting of both sensor and relay nodes is *connected*. The objective of the problem is to minimize $|R|$, where $|R|$ denotes the number of relay nodes in G. Design an approximation algorithm for this minimum relay node placement problem. Comment whether your approximation algorithm works in the scenario where communication radii of the sensor nodes and the relay nodes are different. [10+5=15]

3. (a) Briefly describe the four main types of *device-tier* communications of the 5G cellular networks. [6]

   (b) State the differences between *open access* and *closed access* mechanisms for security provisioning in 5G cellular networks. [4]

   (c) Formulate the optimal resource allocation problem for multiple D2D pairs as a maximum weight bipartite matching problem. [10]

   (d) State two fundamental propagation features of millimeter wave D2D communication in 5G cellular networks. State the fundamental differences between conventional cellular network and millimeter wave cellular network. [4+6=10]

   (e) Briefly describe a model to include the effects of obstacles in a millimeter wave D2D communication network. [5]

INDIAN STATISTICAL INSTITUTE

M. Tech. Computer Science Year II, 2018-19

Semestral Examination

Pattern Recognition and Image Processing

(Back Paper)

Date : 30.01.2019

Total Marks: 100 Maximum Marks: 100 Duration: 3 hours

1.

a. Describe the general principle of a Support Vector Machine, including how it differs from a perceptron classifier. Also give definitions for: [10]
   i. support vector
   ii. margin
   iii. kernel

b. Give one advantage and one disadvantage of Support Vector Machines over other classifiers. [2]

c. The SVM optimization problem L is formulated as a trade-off between two objectives. L can be of the form:

$$L = w_1 A + w_2 B$$

w_1 and w_2 are the weights given to the objectives. What are A and B? Discuss the purpose of both. [8]

2.

a. Describe the working of Principal Component Analysis technique. [10]

b. Why is it not possible for a single binary perceptron to solve the XOR problem? [5]

c. Neural networks require lots of data to be trained properly. If you have too little data (too few input-target pairs) the first thing to try is to get more. However, sometimes this is not possible and, then, to split up the few data you have in a training set and a test set might be considered wasteful. Describe how K-fold cross validation can be used to deal with this problem! [5]

3.

a. Describe mean filter and median filter. Which of them is better and why? [6]

b. Filter the image given below using a 3X3 median filter mask. [10]

24	23	33	25	32	24
34	255	24	0	26	23
23	21	32	32	28	26

c. What is the Sobel operator and for what purpose is it used? [4]

P.T.O.

4.

a. Explain the following operations: [8]
   i. Contrast enhancement
   ii. Bit-plane slicing

b. What is meant by edge linking? Explain edge linking using local processing. [8]

c. What is image compression? Why it is needed? [4]

5.

a. What is image segmentation? Explain any region-based segmentation method along with its drawbacks (if any). [2+8]

d. Describe histogram equalization. Given a 3-bit 4 × 4 image as follows: Perform histogram equalization and find the resulting image. [10]

M=

0	0	0	4
1	1	1	5
1	2	2	7
2	2	2	7

INDIAN STATISTICAL INSTITUTE

Back Paper: 2018-2019

M. Tech. (CS) II year

Data Mining and Knowledge Discovery

Date: ~~12.2018~~ 30.01.2019 **Maximum Marks: 100** **Duration: 3 hours**

1. (i) How is a hierarchical clustering algorithm terminated?
   (ii) Describe algorithm CLARANS.
   (iii) Why do we need density based clustering? [4+8+3=15]

2. (i) Define Minkowski distance?
(ii) Define Jaccard coefficient?
   (v) What are some of the major clustering approaches?
   (vi) What are some of the limitations of k-means clustering?
   (vii) Describe algorithm DBSCAN.

[3+3+4+5+10=25]

3. (i) What are support and confidence in the context of association rules?
(ii) How does FP-tree help in rule mining? Explain tree construction with an example.

[5+10=15]

4. (i) Why can testing out multiple models on your test data be a problem and when is it problematic?
(ii) Describe the dropout method and explain how is it used during training and during inference.
(iii) What is meant by "Batch gradient" descent? What is meant by "stochastic gradient" or "on-line" gradient descent?
(iv) The input image has been converted into a matrix of size 28 X 28 and a kernel/filter of size 7 X 7 with a stride of 1. What will be the size of the convoluted matrix?
(v) In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden output layer and input hidden layer?
(vi) Suppose there is an issue while training a neural network. The training loss/validation loss remains constant. What could be the possible reason?

[2+2+3+4+2+2=15]

5. (i) Derive the gradient descent training rule assuming that the target function representation is: $o_d = w_0 + w_1x_1 + \ldots + w_nx_n$.
(ii) What is regularization, and why do we use it. Give examples of few common methods.

P.T.O.

(iii) Briefly describe overfitting and why it is a problem.
(iv) Briefly describe ways of preventing overfitting in neural networks and why they work.

[5+3+2+5=15]

6. (i) State one advantage of linear rectified activation compared to logistic sigmoid activation.
(ii) Show that the softmax regression model is overparameterized, and in the special case where k = 2 (number of classes) softmax regression reduces to logistic regression.
(iii) Consider that you have a set of digits 2's and 5's from the MNIST dataset. How would you discriminate 2's and 5's using only one autoencoder without any labels?
(iv) How do we choose the size of filter for Convolutional layers? What are the type of activation function for the filters?

[2+5+5+3=15]

Indian Statistical Institute

M. Tech (Computer Science) 2nd year

Mid-Semester Examination - 2019

Advanced Pattern Recognition

Time: **2 hours** Total Marks: **70** Maximum Marks: **60**

Answer the questions briefly

1. Suppose a genetic algorithm uses chromosomes of the form:

$$x = abcdefgh$$

with a fixed length of eight genes. Each gene can be any digit between 0 and 9. Let the fitness of individual x be calculated as:

$$f(x) = (a + b) - (c + d) + (e + f) - (g + h),$$

and let the initial population consists of four individuals with the following chromosomes:

$$x_1 = 6\ 5\ 4\ 1\ 3\ 5\ 3\ 2$$

$$x_2 = 8\ 7\ 1\ 2\ 6\ 6\ 0\ 1$$

$$x_3 = 2\ 3\ 9\ 2\ 1\ 2\ 8\ 5$$

$$x_4 = 4\ 1\ 8\ 5\ 2\ 0\ 9\ 4$$

a) Evaluate the fitness of each individual, showing all your workings, and arrange them in order with the fittest first and the least fit last. (4)

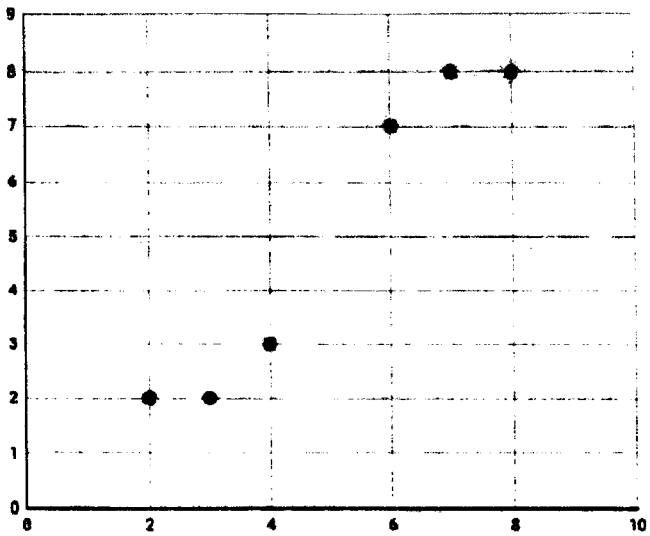
b) Perform the following crossover operations: (4)
   i) Cross the fittest two individuals using one-point crossover at the middle point.
   ii) Cross the second and third fittest individuals using a two-point crossover (points b and f).
   iii) Cross the first and third fittest individuals (ranked 1st and 3rd) using a uniform crossover.

c) By looking at the fitness function and considering that genes can only be digits between 0 and 9 find the chromosome representing the optimal solution (i.e. with the maximum fitness). Find the value of the maximum fitness. (2)

2.
   a. How is GA used to solve the Knapsack problem? (5)
   b. Describe the various properties of fuzzy sets. (5)

P TO

3. With the help of a 6 sample points given below, discuss the Fuzzy C Means clustering algorithm. (10)



4. Discuss the basic structure and the building blocks of a standard CNN. Use appropriate illustrations to aid your explanation. (10)

5.

a. Describe the working principle of a decision tree. (5)
b. What is a random forest? Why are they used? (5)

6.

a. What is shattering? Give examples of 2-D data for n=3, 4 that cannot be shattered. (3)
b. The SVM optimization problem L is formulated as a trade-off between two objectives. L can be of the form:

$$L = w_1A + w_2B$$

w_1 and w_2 are the weights given to the objectives. What are A and B? Discuss the purpose of both. (5)
c. What is the intuition behind the working of a non-linear SVM? (2)

7. Write short notes on: (4X2.5=10)

a. RNN
b. Manifold Learning
c. Transfer learning
d. Explainable AI

X

Midterm
MTech CS Complexity Theory, 2019

19th February, 2019

Time: 2 hours, Maximum Marks: 100

Answer as many questions as possible but the maximum possible marks one can obtain is 100. Your answers should be well-written and you should explain your arguments properly.

1. (20 marks) A language L is called unary if $L \subseteq 1^*$. Show that if a unary language in NP-complete, then $P = NP$.

2. (20 marks) Prove that if every unary language in NP is also in P then $EXP = NEXP$.

3. (20 marks) Show that $DSPACE(n) \neq NP$.

4. (20 marks) Show that there is a language $B \in EXP$ such that $NP^B \neq P^B$.

5. (20 marks) Show that the following language is NL-complete

$$\{G : G \text{ is a strongly connected digraph } \}$$

6. (30 marks) The class DP is defines as the set of languages L for which there are two languages $L_1 \in NP$ and $L_2 \in coNP$ such that $L = L_1 \cap L_2$. Show that

   (a) EXACT-INDSET $\in \Pi_2^p$

   (b) EXACT-INDSET $\in DP$

   (c) Every language in DP is polunomial time reducible to EXACT-INDSET.

   where, EXACT-INDSET is the following language

$$\{(G, k) : \text{ The largest independent set of } G \text{ is of size } k\}$$

Indian Statistical Institute

Computer Vision

M.Tech.(CS)-II 2018-2019

Mid-semestral examination

Full marks: 80 *Date: 20. 2. 2019* Time:3 Hours

Answer **any four** questions. All questions carry equal marks.

1. (a) Consider pin-hole camera model. Prove following two properties of perspective projection:

   (i) Distant objects appear smaller.

   (ii) A set of parallel lines in 3D (not perpendicular to optical axis) maps to a set of concurrent lines in 2D. Hence, define *vanishing point.*

   (b) Consider that the optical centre of digital Camera-1 is at the origin of world coordinate system, and optical axis of the camera coincides with z-axis. The horizontal and vertical edges of the virtual image frame of Camera-1 are parallel to x-axis and y-axis respectively. Suppose we have a second digital camera, namely Camera-2 with the same intrinsic parameters as Camera-1. Camera-2 is placed with its optical axis, and edges of virtual image frame parallel to that of Camera-1, but displaced horizontally with respect to Camera-1 by 6 cm. Two world points P_1 and P_2 are at a distance of 10 cm along z-axis. Both the points are imaged by Camera-1 to same image coordinate (340, 240). However, these points are imaged by Camera-2 to image coordinates (310, 240) and (315, 240). Given that photo receptor density is 100 pixels per cm, calculate the focal length in cm. [(5+(4+1))+10]

2. (a) What is meant by normalized camera? Write down its intrinsic matrix. A normalized camera is placed with its optical centre at $[20 \quad -30 \quad 10]^T$ and rotated anti-clockwise by an angle of 45^o about z-axis. Calculate image coordinate on virtual image plane for a world point (320, 470, 410).

   (b) Consider two normalized digital cameras: Camera-1 and Camera-2. The optical centre of Camera-1 is at the origin, and its optical axis coincides with z-axis. The horizontal and vertical edges of its virtual image frame are parallel to x-axis and y-axis respectively. The pose of Camera-2 may be described by rotation matrix $R_{3\times3}$ and translation vector $T_{3\times1}$ with respect to Camera-1. Derive the expression for *Essential*

PTO

matrix.

(c) An Essential matrix relating points P_1 and P_2 of Image-1 (in Camera-1) and Image-2 (in Camera-2) respectively is given by

$$\begin{bmatrix} 3 & -2 & 4 \\ -2 & 0 & -1 \\ 2 & -3 & 1 \end{bmatrix}$$

Determine epipolar line in Image-2 corresponding to $P_1 = (3, 1, 2)$ and also epipolar line in Image-1 corresponding to $P_2 = (2, 4, 3)$. [5+10+5]

3. (a) Write down the assumptions about noise and signal for *mean filtering*. Hence, derive the expression of mean filter.

(b) i) State two differences between mean and median filters.

ii) Why bilateral filter cleans noise at the edges better than anisotropic diffusion?

iii) What is the underlying strategy of any edge preserving smoothing?

iv) State and explain (why should it work) the basic algorithm for image sharpening.

[(2+2+6)+(2+2+2+4)]

4. (a) i) State the three criteria that are foundation of Canny's edge detector and the major steps in Canny's edge detection algorithm.

ii) Mention how these steps try to satisfy the stated criteria.

(b) Prove

$$\sum_{(x_i,y_i)\in W(x,y)} (f(x_i, y_i) - f(x_i + \Delta x, y_i + \Delta y))^2 = [\Delta x \;\; \Delta y] A_{W(x,y)} [\Delta x \;\; \Delta y]^T$$

where $W(x, y)$ is neighbourhood around (x, y) and $A_{W(x,y)}$ is Harris matrix.

(c) Consider the following Harris matrix

$$\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$$

Given the empirical parameter $k = 0.04$ and $threshold = 10$, calculate corner response by Harris' method and also determine if there exists a corner. [(4+3)+10+3]

5. (a) i) What is the objective of SIFT? What are the major steps of SIFT algorithm?

ii) How does the scale parameter vary within an octave in Lowe's pyramid scheme?

iii) How is first image of an octave obtained from the lower octave?

iv) How is extrema point detected from multi-scale representation?

(b) How is low contrast point eliminated in SIFT algorithm?

(c) Suppose a keypoint is detected, describe how its descriptors is generated from local features. [(2x4)+6+6]

6. (a) i. How can the information processing in visual system be explained in terms of spatial frequency tuned channels?

ii. Is such a theory (as stated in i. above) unique for visual signals only? Justify if you agree, else briefly outline another similar theory in generalized perception.

iii. Briefly describe the emergence of the counter viewpoint to the frequency channel theory?

[3+ (1+3)+ 3]

(b) i. State the Mach's law in vision.

ii. How did Mach's law contribute to the Raw Primal Sketch (RPS) approach to vision?

iii. Hence explain how RPS differs from the earlier approaches to vision.

[3+4+3]

INDIAN STATISTICAL INSTITUTE

M.TECH. (CS) - YEAR II

2018 - 2019

FORMAL METHODS IN COMPUTER SCIENCE – SELECTED TOPICS 20.2.2019

Time: 3 Hours MID-SEMESTER EXAMINATION **Marks: 40**

Answer Question 1 and any 4 from the rest.

1. Let $\mathcal{L}$ be a modal language with 2 modal operators, whose formulas are given as follows:

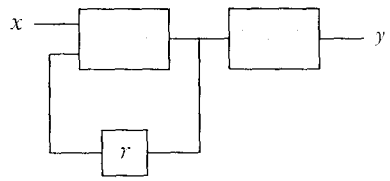
   $$\varphi := \ p \mid \neg\varphi \mid \varphi \to \psi \mid \Diamond_1\varphi \mid \Diamond_2\varphi$$

   where $p \in \mathcal{P}$, a countable set of propositions. Let $\mathcal{M} : (W, R, V)$ be a Kripke model, where W is a non-empty set of worlds, R is a binary relation on W, and $V : \mathcal{P} \to 2^W$ is a valuation function. Let $w \in W$. The truth definition of the formulas at the world w in the model $\mathcal{M}$ is given as follows:

   $\mathcal{M}, w \models p$ iff $w \in V(p)$;
   $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$;
   $\mathcal{M}, w \models \varphi \to \psi$ iff $\mathcal{M}, w \not\models \varphi$ or $\mathcal{M}, w \models \psi$.
   $\mathcal{M}, w \models \Diamond_1\varphi$ iff there exists $w' \in W$ with wRw' such that $\mathcal{M}, w' \models \varphi$
   $\mathcal{M}, w \models \Diamond_2\varphi$ iff there exists $w' \in W$ with $w'Rw$ such that $\mathcal{M}, w' \models \varphi$

   Let φ be a formula in $\mathcal{L}$ such that there exists a model $\mathcal{M}$ and a world w in $\mathcal{M}$ with $\mathcal{M}, w \models \varphi$. Does there exist a finite model $\mathcal{M}'$, say, and a world w' in $\mathcal{M}'$ with $\mathcal{M}', w' \models \varphi$? Justify your answer. [20]

2. Consider the following hardware circuit with input variable x, output variable y and register r. The control function for the output variable y is given by: $y = NOT(XOR(x, r))$. The register evaluation changes according to the circuit function $\delta_r = XOR(x, r)$.



register

   Under the initial register evaluation $[r = 0]$, model the circuit behaviour as a transition system providing justifications. [5]

3. A class of Kripke frames (W, R) with R satisfying a certain property P is said to be defined by a modal formula φ, if for all such frames (W, R), $(W, R) \models \varphi$ iff R satisfies the property P. Here, $(W, R) \models \varphi$ iff $(W, R, V) \models \varphi$ for all valuations V on W. Find the class of frames defined by the following modal operators, with justifications.

   (a) $\Box(\Box\varphi \to \varphi)$
   (b) $\varphi \to \Box\Diamond\varphi$ [5]

4. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two Kripke models, such that $(\mathcal{M}_1, w_1)$ is bisimilar to $(\mathcal{M}_2, w_2)$. Show that for all basic modal formulas φ, $\mathcal{M}_1, w_1 \models \varphi$ iff $\mathcal{M}_2, w_2 \models \varphi$.

   Hence, or, otherwise, show that the unary operator $U\varphi$ defined by:

   $\mathcal{M}, w \models U\varphi$ iff for all $v \in W$, $\mathcal{M}, v \models \varphi$.

5. Let A and B be maximal consistent sets of formulas in basic tense logic. Show that the following are equivalent:

   (a) whenever $\varphi \in A$, we have $P\varphi \in B$

   (b) whenever $\psi \in B$, we have $F\psi \in A$

   (c) whenever $G\sigma \in A$, we have $\sigma \in B$

   (d) whenever $H\tau \in B$, we have $\tau \in A$ [5]

6. Let A, B and C be maximal consistent sets of formulas in the tense logic of rational numbers. Two such maximal consistent sets M and N are said to related, denoted by $M \mathcal{R} N$, if for all formulas φ, $\varphi \in N$ implies $F\varphi \in M$. Prove the following statements:

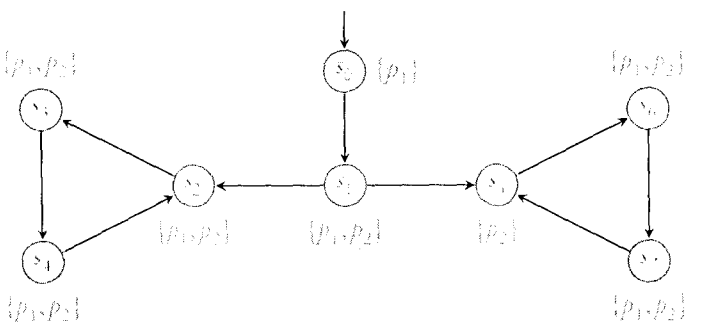
   (a) If $A \mathcal{R} B$ then there exists a maximal consistent set D, such that $A \mathcal{R} D$ and $D \mathcal{R} B$.

   (b) If $A \mathcal{R} B$ and $A \mathcal{R} C$, then either $B = C$ or $B \mathcal{R} C$ or $C \mathcal{R} B$. [5]

7. Provide model checking algorithms for the CTL formulas of the form $EX\varphi$, $EG\varphi$ and $E(\varphi U \psi)$. Using the algorithms, find the worlds in the model below where the following formulas hold: [5]

   (a) $EXEG(p_1 \wedge p_2)$

   (b) $E(p_2 U p_1)$



Indian Statistical Institute
Mid-Semester Examination: 2019
Course Name: M. Tech in Computer Science
Subject Name: Computer Networks

Date: 21-02-2019 Maximum Marks: 60 Duration: 3 hours

Instructions:

You **may** attempt **all** questions which carry a total of **65** marks. However, the maximum marks you can score is only **60**.

1. (a) For the bit stream 101011100, sketch the waveforms for each of the following encoding schemes.
   i. NRZ-L
   ii. Differential Manchester. [3+3=6]
   (b) State three design goals of the scrambling technique? [3]
   (c) If an optical fiber has a bandwidth of 2 GHz and a modem uses 512 signal levels, what is the maximum data rate according to Nyquist? If the average signal power is 405 units and the average noise power is 27 units, what is the maximum channel capacity according to Shannon? [3+3=6]
   (d) Consider a digital telephone system which uses Pulse Code Modulation (PCM) as encoding scheme. Assuming the human voice has a spectrum of frequencies ranging from 200 Hz to 4000 Hz, what sampling rate should be used to retain all necessary information of the original voice signal? Assume that the number of different code levels used for PCM in the said telephone system is 128. If ISI Kolkata has a 1 Mbps dedicated link to ISI Bangalore, how many PCM encoded voice calls can be sent from ISI Kolkata to ISI Bangalore at the same time (ignore other overheads such as headers)? [3+3=6]
   (e) Knowledge of the quantization level (δ) alone is insufficient to make a statement about the possibility of slope overload noise in delta modulation. What else needs to be known? [1]

2. (a) What is a pseudo noise (PN) sequence? State the autocorrelation property of a maximum length PN sequence? [3+3=6]
   (b) Briefly describe the basic principle of direct-sequence spread spectrum. [5]
   (c) Consider an MFSK scheme with carrier frequency f_c equal to 250 KHz, difference frequency f_d equal to 25 KHz, number of different signal elements M equal to 8, and number of bits per signal element L equal to 3.

   Suppose we wish to apply frequency hopping spread spectrum (FHSS) to this MFSK scheme with $k = 2$; that is, the system will hop among 4 different carrier frequencies. Let T_c be the period at which the MFSK carrier frequency changes and T_s is the duration of a signal element. Consider a **fast FHSS** with T_s being $4T_c$. Show the sequence of frequencies used, and the times the frequency changes occur, for transmitting the bit string 011110001. Assume that the PN sequence is 0011 0111 0010 1101 0001 1111. [5]
   (d) A bit string, 0111101111101111110, needs to be transmitted at the data link layer. What is the string actually transmitted after bit-stuffing? [2]

(e) Given the output after byte-stuffing: FLAG A B ESC ESC C ESC ESC ESC FLAG ESC FLAG D FLAG. What is the original data? [2]

3. (a) Consider the use of CRC with generator polynomial $G(x) = x^4 + x^3 + 1$ for error detection. Show the transmitted string for the message string 1 1 0 0 1 1. Construct a burst error of length 5 on the transmitted string in such a way that the error cannot be detected by the CRC with the given $G(x)$. [3+3=6]

(b) Consider a repetition code where each bit b is encoded as n copies of b. Assume that the maximum likelihood decoding rule is employed at the receiver. Assuming binary symmetric channel (BSC) model, compute the probability of decoding error when bit error rate is ϵ and n is an even integer. [3]

(c) Consider a linear block code over $\mathbb{F}_2$ with three data bits D_1, D_2, D_3 and three parity bits P_1, P_2, P_3. The parity bits are defined as follows:

$$P_1 = D_1 + D_2$$
$$P_2 = D_2 + D_3$$
$$P_3 = D_3 + D_1.$$

i. Find the minimum Hamming distance of the code. [3]

ii. What are the error detection and error correction capabilities of the code? [1+1=2]

iii. Suppose the receiver computes three syndrome bits E_1, E_2, E_3 from the (possibly corrupted) received data and parity bits as follows:

$$E_1 = D_1 + D_2 + P_1$$
$$E_2 = D_2 + D_3 + P_2$$
$$E_3 = D_3 + D_1 + P_3.$$

Suppose the receiver performs maximum likelihood decoding using the syndrome bits. Consider the following four combinations of syndrome bits: (i) $E_3 = 0, E_2 = 0, E_1 = 0$, (ii) $E_3 = 0, E_2 = 1, E_1 = 0$, (iii) $E_3 = 1, E_2 = 0, E_1 = 1$, *and* (iv) $E_3 = 1, E_2 = 1, E_1 = 1$. For each combination of syndrome bits, state what the maximum-likelihood decoder believes has occurred: no errors, a single error in a specific bit (state which one), or multiple errors. [4]

(d) Recall that a block code takes a set of k-bit messages and produces n-bit codewords, with a minimum Hamming distance of d between any two codewords. State whether a linear block code with parameters $(n, k, d) = (32, 27, 3)$ exists or not. If such a code exists, give an example; if not, explain why not. [3]

(e) Consider the set of codewords {00000, 01111, 10100, 11011} of a code and state whether the code is a linear block code over $\mathbb{F}_2$ or not. [2]

(b) about 3-5 times lesser storage than the corpus.

(c) about 50-70 times lesser storage than the corpus.

(d) about 3-5 times more storage than the corpus.

(e) may require more, equal, or lesser storage than the corpus.

4. Given a distributed file containing 10 million integers, where every node containing some multiple of 10000 integers, a MapReduce program is defined by the following map and reduce functions:

   map: For every batch of 10,000 integers, emits key value pair $< 1, m >$, where m is the mean of the integers.

   reduce: For each key k, output the mean of the values.

   The program would output

   (a) The median of the integers in the file

   (b) The (arithmetic) mean of the integers in the file

   (c) The harmonic mean of the integers in the file

   (d) The geometric mean of the integers in the file

   (e) None of the other answers is necessarily correct

5. If PageRank is computed without teleporting, and if there are dead ends (nodes with no outlink) in a graph, then

   (a) The PageRank for all nodes become zero.

   (b) The PageRank remains positive for the dead ends, and zero for all other nodes.

   (c) The PageRank becomes zero for the dead ends and remain positive for all other nodes.

   (d) The total PageRank becomes more than 1.

   (e) The PageRank becomes 1 for each of the dead ends, and zero for all other nodes

6. Suppose $\boldsymbol{M} = (m_{ij})_{i,j=1}^{n}$ is an $n \times n$ matrix and $\boldsymbol{v} = (v_k)_{k=1}^{n}$ is a vector of size n. A MapReduce algorithm is defined by the following map and reduce functions.

   map: For every non-zero entry m_{ij} of $\boldsymbol{M}$, emit the key value pair $< j, m_{ij}v_i >$.

   reduce: For every key j, output $\sum_{i=1}^{n} m_{ij}v_i$ as the j-th entry of a vector.

   The above MapReduce algorithm computes

   (a) $\boldsymbol{v}^T\boldsymbol{M}$

   (b) $\boldsymbol{M}\boldsymbol{v}$

   (c) $\boldsymbol{M}^2\boldsymbol{v}$

   (d) $\boldsymbol{M}\boldsymbol{v}^2$

   (e) None of the other answers

Information Retrieval

Mid Sem Examination
M.Tech CS 2nd Year, 2nd Semester
Full marks: 60
Time: 120 minutes

21 February 2019

Instructions:

- Use blank answersheets for rough. Attach this question paper along with the rough sheet and submit.
- Each question has at least one correct answer. Tick a correct answer. A correct answer will get 4 marks, a wrong answer will get -1, and if you do not attempt a question, you will get zero. If you tick multiple answers, it will be considered as not attempting the question.
- This examination is **not** open notes, or books. Use of no digital device is permitted.

1. Suppose the unary representation of 7 is 11111110 (7 ones, followed by a zero). Then, the Elias-γ code for 31 is
   (a) 1111101111
   (b) 111101111
   (c) 000011111
   (d) 11111100111
   (e) None of the other answers is correct
2. More skip pointers in posting lists are expected to
   (a) increase comparison overhead, but more chances of skips
   (b) decrease comparison overhead, but less chances of skips
   (c) increase comparison overhead and less chances of skips
   (d) decrease comparison overhead and more chances of skips
   (e) none of the other answers are correct
3. An uncompressed inverted index built from a 1TB uncompressed text corpus (assuming the index stores, for every term, the list od document ids and tf.idf scores, without position information) is expected to consume
   (a) about the same amount of storage space as the corpus.

(b) about 3-5 times lesser storage than the corpus.

(c) about 50-70 times lesser storage than the corpus.

(d) about 3-5 times more storage than the corpus.

(e) may require more, equal, or lesser storage than the corpus.

4. Given a distributed file containing 10 million integers, where every node containing some multiple of 10000 integers, a MapReduce program is defined by the following map and reduce functions:

   map: For every batch of 10,000 integers, emits key value pair $< 1, m >$, where m is the mean of the integers.

   reduce: For each key k, output the mean of the values.

   The program would output

   (a) The median of the integers in the file

   (b) The (arithmetic) mean of the integers in the file

   (c) The harmonic mean of the integers in the file

   (d) The geometric mean of the integers in the file

   (e) None of the other answers is necessarily correct

5. If PageRank is computed without teleporting, and if there are dead ends (nodes with no outlink) in a graph, then

   (a) The PageRank for all nodes become zero.

   (b) The PageRank remains positive for the dead ends, and zero for all other nodes.

   (c) The PageRank becomes zero for the dead ends and remain positive for all other nodes.

   (d) The total PageRank becomes more than 1.

   (e) The PageRank becomes 1 for each of the dead ends, and zero for all other nodes

6. Suppose $\boldsymbol{M} = (m_{ij})_{i,j=1}^{n}$ is an $n \times n$ matrix and $\boldsymbol{v} = (v_k)_{k=1}^{n}$ is a vector of size n. A MapReduce algorithm is defined by the following map and reduce functions.

   map: For every non-zero entry m_{ij} of $\boldsymbol{M}$, emit the key value pair $< j, m_{ij}v_i >$.

   reduce: For every key j, output $\sum_{i=1}^{n} m_{ij}v_i$ as the j-th entry of a vector.

   The above MapReduce algorithm computes

   (a) $\boldsymbol{v}^T\boldsymbol{M}$

   (b) $\boldsymbol{M}\boldsymbol{v}$

   (c) $\boldsymbol{M}^2\boldsymbol{v}$

   (d) $\boldsymbol{M}\boldsymbol{v}^2$

   (e) None of the other answers

7. Recall Rocchio's iterative formula for refining queries in relevance feedback framework:

$$q_m = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{d \in D_r} d - \gamma \frac{1}{|D_{nr}|} \sum_{d \in D_{nr}} d$$

where q_0 is the original query. The set D_r is

(a) ideally the set of all relevant documents, but approximated by the whole corpus.

(b) ideally the set of all relevant documents, but approximated by the set of known relevant documents.

(c) ideally the set of known relevant documents, but approximated by the whole corpus.

(d) the set of top 10 documents retrieved using q_0.

(e) None of the other answers is correct.

8. If the posting lists of an inverted index are ordered by decreasing tf.idf scores, then

(a) Merge join can be performed, but not NRA

(b) NRA algorithm can be executed, but not merge join

(c) Both NRA algorithm and merge join may be performed

(d) None of NRA algorithm and merge join may be performed

(e) Only if the lists contain authority (such as PageRank) scores together with tf.idf scores, then merge join can be performed

9. Computing the intersection of two posting lists of length n each takes (select the best possible complexity)

(a) $O(n^2)$ time

(b) $O(n)$ time

(c) $O(n \log n)$ time

(d) $O(\sqrt{n} \log n)$ time

(e) None of the other answers

10. Finding the top-k numbers from a list of n unsorted numbers can be done in (select the best possible complexity)

(a) $O(n \log n)$ time

(b) $O(n + k \log n)$ time

(c) $O(k + n \log k)$ time

(d) $O(n \log k)$ time

(e) None of the other answers

11. A variable byte (with 1 byte units) gap encoding for the posting list $< 24, 48, 100, 240 >$ would consume

(a) 4 bytes

(b) 5 bytes

(c) 7 bytes

(d) 8 bytes

(e) 16 bytes

12. Suppose a precision-recall plot for a single query (with recall on the x-axis and precision on the y-axis) is parallel to the x-axis, with precision = 1 all the way. Then which of the following is always true?

    (a) The first and second documents are relevant, the third one is not.

    (b) Every relevant document is ranked higher than every non-relevant document.

    (c) Every non relevant document is ranked higher than every relevant document.

    (d) The first and third documents are relevant, but the second one is not.

    (e) None of the other answers are necessarily correct.

13. On a given query, if the mean reciprocal rank for a retrieval system is 1/3, then

    (a) All of the first, second and third documents is the correct answer

    (b) Any one of the first, second and third documents can be the correct answer

    (c) The second document is the correct answer, but the first and third are not

    (d) The third document is the correct answer, but the first and second are not

    (e) None of the first three document is the correct answer

14. When PageRank for all nodes in a graph G with $n > 5$ nodes is computed with teleporting, using the following iterative formula $\boldsymbol{v}_{k+1} = \beta \boldsymbol{M} \boldsymbol{v}_k + (1-\beta)\frac{1}{n}$, then the minimum PageRank for a node in G would be

    (a) β/n

    (b) $(1-\beta)/n$

    (c) $1/n$

    (d) 0

    (e) None of the other answers is necessarily correct

15. Apart from the query, what input does pseudo relevance feedback require from the user?

    (a) Input on a few relevant documents only

    (b) Input on exactly one document that the user considers relevant

    (c) No further input from the user

    (d) Input on a few relevant and non-relevant documents

    (e) Input on a few non-relevant documents only

INDIAN STATISTICAL INSTITUTE

Periodical Examination

M. Tech (CS) - II Year (Semester - II)

Combinatorial Geometry

Date : 21.02.2019 Maximum Marks : 60 Duration : 3 Hours

Note : You may answer any part of any question, but maximum you can score is 60.

1. Given a set $P = \{p_1, p_2, \ldots, p_n\}$ of n points in $\mathbb{R}^d$, define $f(P)$ as the number of pairs of points $p_i, p_j \in P$, $i < j$, such that the distance $\delta(p_i, p_j)$ between p_i, p_j is equal to 1. Also define $f_d(n)$ as the $\max_{P \subseteq \mathbb{R}^d, |P|=n} f(P)$, in other words, the maximum number of unit distances determined by any point set P of size n.

   Show that $f_1(n) = n - 1$, and $f_2(n) \leq cn^{\frac{3}{2}}$, where c is a known positive constant.

   [2+10=12]

2.(a) Consider the triangulation of a point set of size n ($n \geq 3$) with k edges on its outer face. Show that the number of edges in the triangulation is $3n - 3 - k$.

(b) Show that, if G is a graph with n (≥ 3) vertices and m ($\geq 4n$) edges then the crossing number of the graph G in its planar embedding satisfies $Cr(G) \geq m - 3n + 6$. [4+6=10]

3.(a) Let L be a set of n lines in the plane and let r be a parameter $1 < r < n$.
Show that there a partitioning of the plane into t generalized triangles $\Delta_1, \Delta_2, \ldots, \Delta_t$ such that each Δ_i is intersected by at most $\frac{n}{r}$ lines of L, and $t = O(r^2 \log^2 n)$.

(b) Show that given any set of n non-collinear points in the plane determines at least n different connecting lines, i.e., lines through at least two points of the set.

(c) Also show that n points define exactly n connecting lines if and only if all but one of the points are collinear.

[10+10+5=25]

4.(a) For a pair of integers n, s, define the Davenport Schinzel sequence $DS(n, s)$.
If $f_1, f_2, \ldots, f_n$ are functions such that $f_i \neq f_j$ for $i \neq j$, and every pair of functions f_i, f_j intersects at most s times, then show that the indices of the segments of $f_1, f_2, \ldots, f_n$ that appears in the lower envelope of $f_1, f_2, \ldots, f_n$ is a $DS(n, s)$ sequence.

(b) Also show that, given a $DS(n, s)$ sequence, one can construct a set of functions whose lower envelope corresponds to the given $DS(n, s)$ sequence.

(c) Write an algorithm to compute the lower envelope of a set of functions $f_1, f_2, \ldots, f_n$ where every pair of functions f_i, f_j intersects at most twice.

[4+8+10=22]

5. Define halfspace centerpoint in $\mathbb{R}^2$.
   Show that for a finite point set $X \in \mathbb{R}^2$, there always exists a a halfspace-centerpoint.
   Given a point set $X \in \mathbb{R}^2$, $|X| = n$, write an algorithm to compute a half-space centerpoint of X.
   State and justify the time complexity of your proposed algorithm.

[3+6+6+5=20]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination 2018-2019 — M. Tech. (CS) Second Year

Subject: VLSI Testing and Fault tolerance — Date: 22.2.2019

Maximum marks = 60 — Time: 2 hours

Instructions: *Answer all questions. Answer parts of a question in the same place*

---

1. For each of the following statements, indicate whether the statement is true or false. For the ones that are true, provide a brief justification. For the ones that are false, provide a small counter-example. **[8 x 2 = 16 marks]**
   (a) All LTL properties can be expressed in CTL.
   (b) In the *counter-example guided abstraction refinement* (CEGAR) approach, the abstract machine has fewer runs as compared to the original machine.
   (c) The complexity of both LTL model checking and CTL model checking are linear in the size of the Kripke structure.
   (d) A bounded model checking tool needs to add clauses for loop detection while searching for counter-example traces of LTL properties of the form $\varphi \ \mathbf{U} \ \psi$.
   (e) Any CNF formula in which there is not more than one positive literal in each clause, is solvable in polynomial time.
   (f) There exist languages, which are accepted by some non-deterministic Buchi automaton but not by any deterministic Buchi-automaton.
   (g) The worst case space complexity of a DPLL SAT solver without learning is polynomial in the size of the original problem.
   (h) The size of the ROBDD of $(a1 \wedge b1) \oplus (a2 \wedge b2)$ for the variable ordering $(a1 < a2 < b1 < b2)$ is less than the one generated for the same function for the ordering $(a1 < b1 < a2 < b2)$.

2. [Model Checking] **[5 + 7 + 4 + 4= 20 marks]**
   (a) Let AP = {a} and $\varphi = (a \wedge X\, a)\ \mathbf{U}\ \neg\, a$ be an LTL formula over AP.
      (i) Is φ satisfiable? If yes, give an example run. If not, justify your answer.
      (ii) Construct the Büchi automaton B_φ such that $L\omega(B_\varphi) = \text{Words}(\varphi)$.
      (iii) Explain your satisfiability result above on this automaton.
   (b) The language ACTL restricts CTL to using only the universal ($\forall$) quantifier. Is ACTL same as LTL? Either prove that these are equivalent, or present examples to show that they differ in expressibility.

3. [Abstractions and Refinement] **[4 + 6 + 4 = 14 marks]**

(a) If two states of a finite state machine are bi-simulation equivalent, then do they satisfy the same set of LTL properties? Is the converse true? Justify both answers.

(b) Draw a small FSM having 3 state variables, and its abstraction with 2 visible and 1 invisible variables. Define a LTL property over the visible state variables for which the abstract state machine produces a counter-example trace, which cannot be reproduced in the original FSM.

(c) How does the *counter-example guided abstraction refinement* (CEGAR) approach handle counter-examples like the one in part (b)?

4. [Specification Design] **[3 + 3 + 4 = 10 marks]**

A Gray counter has the property that successive values of the counter differ in only one bit. The exact counting sequence may vary from implementation to implementation. For example, one valid counting sequence for a 3-bit Gray counter is:

000 → 001 → 011 → 010 → 110 → 111 → 101 → 100 → 000 ...

Another valid counting sequence for a 3-bit Gray counter is:

000 → 010 → 110 → 100 → 101 → 111 → 011 → 001 → 000 ...

Let **s [0:2]** denote the 3-bit state vector of the counter.

(i) Write the following properties in Linear Temporal Logic (LTL):

- Successive states of the counter differ in exactly one bit.
- A visited state is re-visited every 8 clock cycles.

(ii) Do we need any more properties for specifying a 3-bit Gray counter? If so, add more properties. If not, justify that any implementation, which satisfies these two properties is a valid 3-bit Gray counter.

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2019

Subject: **Advanced Cryptology** Course Name : M.Tech. (CS) II yr

Max Score: 35 Date: 22 Feb. Duration: 120 Mins

Note: Attempt all questions. Marks are given in brackets. Total score is 40. In $O(\cdot)$ notation we ignore the log factor.

1 [8] Let $\Delta(X;Y)$ denote the statistical distance between two random variables X and Y. Suppose $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are n independent copies of X and Y respectively. Then show that

$$\Delta((X_1, \ldots, X_n); (Y_1, \ldots, Y_n)) \leq n\Delta(X;Y).$$

2 [8] Let $g : \{0,1\}^n \to \{0,1\}^{2n}$ and A be an efficient preimage finder algorithm. For $y \in \{0,1\}^{2n}$ we say that $A(y)$ wins if it returns x such that $g(x) = y$. Let p denote the probability that A wins with the input $g(U)$ where U is the uniformly selected from $\{0,1\}^n$. Construct a PRBG adversary B against g such that

$$\mathbf{Adv}_g^{\text{prbg}}(B) \geq p - \frac{1}{2^n}.$$

3 [8] Let $g : \{0,1\}^n \to \{0,1\}^n$ be a bijective function. We define a compression function $G : \{0,1\}^{2n} \to \{0,1\}^n$ as $G(x,y) = g(x) \oplus g(y)$. Write down a collision finding algorithm for G which runs in time and memory $O(2^{n/3})$.

4 [3+5=8] Define dual sponge Authenticated Encryption (AE) algorithm with rate 100 and capacity 100. Construct a privacy attack against dual sponge AE with time complexity 2^{80} and data complexity 2^{20}.

5 [8] Let $P : \{0,1\}^n \to \{0,1\}^n$ be a permutation. Let $E_{K_1,K_2}(x) : \{0,1\}^n \to \{0,1\}^n$ be defined as $P(x \oplus K_1) \oplus K_2$ for two n-bit keys K_1, K_2. Find a key-recovery algorithm with complexity $O(2^{n/2})$.

Indian Statistical Institute
Information Retrieval
M.Tech. CS 2nd Year 2nd Semester
Final Examination 2018-19

Time: 3 hours
Total marks: 100

April 18, 2019

Part A: Answer any 5 questions from Questions 1-6. Each question carries 12 marks.

1. Refer to the document collection below and draw an inverted index for the collection with positional information.

   doc 1: Election in India next month.
   doc 2: Political campaigns all over India.
   doc 3: Indian Premier League schedule adjusted due to election.
   doc 4: Election results may affect stock market.

2. Explain, with example if necessary, whether each of the following statements is true:

   (a) "Stemming does not lower recall in case of Boolean retrieval."

   (b) "Stemming does not lower precision in case of Boolean retrieval."

3. Describe a data structure that can be used to process *leading* wildcard queries, i.e., queries of the form `*ata`. Explain how it works using the query `*ata` as an example.

4. Consider the following question: *"What is the capital of India?"*. Describe a simple query rewriting algorithm to answer questions of this type using a large web-corpus. In particular, explain how it would answer the given question.

5. Apart from *term-spam*, describe two methods with brief explanation for spamming the web.

6. After a retrieval system computes the list of n documents (for some large n) and their scores for a query by merging multiple posting lists which are sorted by document id, it needs to compute the k (for some small k) documents with the highest score to be able to return the top-k. Briefly describe an efficient algorithm for finding the top-k from the large list of n documents with scores and derive the complexity of your algorithm.

Part B: Answer any 2 questions from Questions 7-10. Each question carries 20 marks.

7. Let $A = (a_{ij})$ be the Boolean term-document matrix representing a document collection, i.e., $a_{ij} = 1$ if term-i occurs in document-j, 0 otherwise. Let $A = U\Sigma V^T$ define the singular value decomposition of A. Suppose you have access to only the matrices U and Σ (not A or V). Describe a method to compute the number of documents in which term-i and term-j occur together, using only U and Σ.

8. A social network supports tagging users in photos and stores the data in a HDFS. The tags given to each photo are stored as a list. An example is given below.

```
photo 1: user1, user2, ...
photo 2: user4, user7, ...
```

Since the number of photos is very large, the list of tags for one photo may be assumed to be in a single node, but the lists for different photos may be in different nodes.

Write the *map* and the *reduce* functions for a MapReduce method which would compute all pairs of users who are tagged together in at least two photos.

9. Consider a web-advertising scenario with two bidders P and Q, each with budget $B = \$8$, and each bid being of \$1. Describe a bidding scenario and a query stream for which the greedy algorithm generates more revenue (by using up more budget of the bidders) than the balance algorithm, or prove that it is not possible. Note that ties are broken arbitrarily for both of the algorithms. Hence, for constructing a special scenario, you can break ties as per your choice.

10. Let X be a set of bit vectors of size n. Let the Hamming similarity $HS(x, y)$ between two bit vectors $x, y \in X$ be defined as m_{xy}/n, where m_{xy} is the number of bits in which x and y have the same value. Let $h : X \to \{0, 1\}$ be a function which projects a bit vector x to any of its bits randomly (in other words, $h(x)$ is the i-th bit of x, where i is chosen randomly).

Prove that h has the LSH property, i.e., for any two bit vectors $x, y \in X$,

$$P[h(x) = h(y)] = HS(x, y).$$

Endterm
MTech CS Complexity Theory, 2019

18th April, 2019

Time: 3 hours, Maximum Marks: 100

Answer as many questions as possible but the maximum possible marks one can obtain is 100. Your answers should be well-written and you should explain your arguments properly.

1. (15 (5+5+5) marks) Which of the following languages are in NP, which one in co-NP, which of them are in P and which one is NP-complete. Give explanation for your answers.
   (a) $IS_k = \{G \mid$ Graph G has an independent set of size $k\}$.
   (b) $MIS_k = \{G \mid$ Maximum independent set of graph G is of size exactly $k\}$.
   (c) $MaxMatching_k = \{G \mid$ Maximum matching of graph G is of size exactly $k\}$.
2. (15 (5+5 +5) marks) Suppose $L_1, L_2 \in NP$, then which of the following is true (explain)
   (a) $L_1 \cap L_2 \in NP$.
   (b) $\overline{L_1} \in NP$ (where, $\overline{L_1}$ is the complement of L_1).
   (c) If L_1 and L_2 is also NP-complete, then is $L_1 \cap L_2$ also NP-complete?
3. (15 marks) Show that there exists a function that is not space-constructable.
4. (15 marks) Show that 2SAT is in NL.
5. (15 marks) Describe a language in P/poly that is not in P.
6. (15 marks) Show that BPL $\subseteq$ P. (Recal that the class BPL is the set of languages for which there is a bounded error probabilistic logspace algorithm).
7. (15 marks) Describe an algorithm for Polynomial Identity Testing in BPP. (Hint: Use Schwartz-Zippel Lemma)
8. (10 (2+2+2+2+2) + 15 (5+5+5) marks) Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Explain the following terms
   (a) sensitivity of f (i.e. $s(f)$)
   (b) block sensitivity of f (i.e. $bs(f)$)

(c) certificate complexity of f (i.e. $C(f)$)

(d) decision tree complexity of f (i.e. $D(f)$)

(e) randomized decision tree complexity of f (i.e. $R(f)$)

Let f be a function on $n = k^2$ variables that is the **AND** of k **OR**'s, where each **OR** is on a disjoint set k variables. Show that

(a) $s(f) = bs(f) = C(f) = n$

(b) $D(f) = n$

(c) $R(f) \geq \Omega(n)$

Indian Statistical Institute

Computer Vision

M.Tech.(CS)-II 2018-2019

Semestral examination

Full marks: 100 *Date: 22. 4. 2019* Time:3 Hours

Answer **any ten** questions. All questions carry equal marks.

1. (a) Using inverse perspective transformation or otherwise, show that depth information is imperative to calculate the world coordinates from camera coordinates.

   (b) Draw a neat diagram outlining the following configuration: two cameras (Left and Right), with focal length f for each, are kept separated by a distance Δx along the x-axis. Two image points P_L $(x_L,\ y_L)$ and P_R $(x_R,\ y_R)$ are formed on the Left and Right cameras, respectively, corresponding to a scene point P $(x_o,\ y_o,\ z_o)$. Now consider all the positions and distances mentioned above to be measured from the origin of the left-hand cameras co-ordinate system, and compute the depth information z_o for the scene point P in terms of f, Δx and D only, where $D = x_L - (x_R + \Delta x)$. [3+(2+5)]

2. (a) Define Binocular Disparity (BD) in the Lateral Stereoscopic Vision model, clearly explaining all the terms involved. Explain diagrammatically what happens when $BD \to 0$. Hence explain the significance of the terminology Binocular Disparity.

   (b) Do you agree to BD being referred to as Relative Shift? Justify your answer. [(2+3+2)+3]

3. (a) Consider pin-hole camera model. Prove following property of perspective projection: A straight line in 3D maps to a straight line in 2D.

   (b) Consider that the optical centre of a digital camera is at the origin of world co-ordinate system, and optical axis of the camera coincides with z-axis. The horizontal and vertical edges of the image frame of the camera are parallel to x-axis and y-axis respectively. Distance from the optical centre to image plane is 5cm, size of image frame is 3cm $\times$4cm, and photo receptor density is 1200 pixels per cm along vertical direction and that along horizontal direction is 1600 pixels per cm. Assuming pin-hole camera model, determine the image coordinate for a world point located at $(X,\ Y,\ Z)$ $= (6cm,\ 8cm,\ 15cm)$. [5+5]

P.T.O.

4. Consider that the optical centre of a digital camera is at the origin of world co-ordinate system, and optical axis of the camera coincides with z-axis. The horizontal and vertical edges of the virtual image frame of the camera are parallel to x-axis and y-axis respectively. Consider that photo receptor density is 100 pixels per cm and the focal length is 10 cm. Suppose a world point is imaged at (340, 240) in pixels. If the camera is shifted horizontally by 6 cm, same world point would be imaged at (310, 240) in pixels.

   (a) Write down the intrinsic matrix of the camera.

   (b) Calculate the depth of the world point in cm. [5+5]

5. (a) Suppose there are two digital cameras where second camera is rotated about z-axis and y-axis both by 45^o and shifted by a translation vector $(3 \ -2 \ 1)$. Determine the essential matrix relating the image points between them.

   (b) Derive the expression for Fundamental Matrix relating image points of two general cameras. [3+7]

6. (a) Why is expression $\rho = x\cos\theta + y\sin\theta$ used instead of $y = mx + c$ for detecting lines by Hough transform?

   (b) Suppose in computing HoG features, the orientation range 180^o is divided into 9 bins as 0-20, 20-40, etc. How would gradient magnitude 120 unit with orientation 55^o be distributed between two nearest bins following linear voting?

   (c) In SIFT key point detection, at what condition $\frac{trace(H)}{det(H)}$ be minimum and what is that minimum value? H denotes Hessian matrix. [2+2+6]

7. (a) What is meant by *optical flow* and *motion field*? Give one difference between them.

   (b) Prove that the magnitude of motion vector of the optical flow is inversely proportional to the depth of the object. [2+3+5]

8. (a) Subject to brightness constancy constraint, show that the tempral gradient is related to motion vector.

   (b) Consider 5 patches of 5 frames (with indices t-2, t-1, t, t+1 and t+2) correspondng to same spatial location as follows:

P.T.O.

50	50	40	40	50
40	50	50	50	40
50	40	40	50	40
50	50	50	40	50
50	40	50	40	40

40	50	30	40	30
30	40	40	30	50
40	30	50	40	30
50	40	30	50	40
30	50	30	50	50

40	30	40	40	30
30	40	30	30	40
40	30	40	30	30
30	40	30	40	30
30	40	30	40	40

30	40	30	30	40
40	30	40	40	40
30	40	30	40	40
30	40	30	40	30
40	30	40	30	30

30	20	30	30	20
20	30	20	30	20
30	20	30	20	30
30	20	30	20	20
20	30	20	30	30

Using Simoncelli's algorithm (1994), calculate temporal gradient I_t at central location (x, y, t), where 1-D convolution kernel is given by

index	-2	-1	0	1	2
p_5	0.1	0.2	0.4	0.2	0.1
d_5	-0.1	-0.3	0.0	0.3	0.1

[(2+3)+5]

9. (a) Write down the objective function proposed by Lucas and Kanade for detecting optical flow, and hence derive the expression for optical flow vector.

   (b) Write down the steps of Horn and Schunck iterative algorithm for computing optical flow. [(2+3)+5]

10. (a) Define (i, j)-th central moment of an image. Show that the central moment is a measure of symmetry, when $(i + j)$ is odd.

    (b) Define gray level co-occurrence matrix.

    (c) Consider the following 8×8 image. Compute the graylevel co-occurrence matrix for

P.T.O.

distance between pixels equal to one and direction is horizontally right.

0	2	2	2	3	3	1	0
0	3	1	2	0	2	1	2
2	1	0	2	2	1	3	2
3	3	0	0	2	3	2	3
3	2	0	1	3	2	2	1
0	0	2	3	1	2	3	2
0	1	2	1	1	2	3	3
0	2	3	2	3	1	0	2

[(1+4)+(1+4)]

11. (a) How orientation assignment is done in SURF after getting the interest points? How do we form the descriptors in SURF?

    (b) Given a three views of a scene and a pair of matched points in two views, write down the steps to determine the position of the point in the third view. [(2+3)+5]

12. (a) What are the major drawbacks of meanshift? How does CAMSHIFT address these issues?

    (b) What is Texture Aliasing? Why does it occur? Explain the basic difference between Bump Mapping and Normal Mapping. [(2+3)+(1+2+2)]

13. (a) Write step by step Grabcut Algorithm for foreground extraction. Mention briefly two approaches to solve energy minimization problem.

    (b) State the problem(s) of binary patterns in achieving high spatial resolution in estimating depth using structured light. How does gray level pattern solve this problem? [(2+3)+5]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

M. Tech (CS) - II Year (Semester - II)

Combinatorial Geometry

Date : 22.04.2019 Maximum Marks : 100 Duration : 3 Hours

Note : You may answer any part of any question, but maximum you can score is 100.

1. For every finite point set $P \in \mathbb{R}^2$ in general position, show that there exists a point $\pi \in \mathbb{R}^2$ (π may not be in P) such that a pair of mutually intersecting lines through π divides P into four quadrants where each quadrant contains at least $\lfloor \frac{|P|}{4} \rfloor$ points of P. [10]

2. Prove the *zone theorem*, stated below:

   Given an arrangement $\mathcal{A}(L)$ of a set L of n lines, and a line $\ell \notin L$, the number of edges of $\mathcal{A}(L)$ intersecting ℓ is at most $6n$.

   [10]

3.(a) *Halfspace centerpoint* in $\mathbb{R}^d$ is defined below:

   A *halfspace centerpoint* of a set P of n discrete point set in $\mathbb{R}^d$ is a point $\pi \in \mathbb{R}^d$ such that each closed halfspace that contains π also contains at least $\frac{n}{d+1}$ points of P.

   (b) Show that for every finite point set $P \subseteq \mathbb{R}^d$, there exists a half-space center-point π.

   (c) Define an α-quadrant centerpoint of a point set $P \in \mathbb{R}^2$.

   (d) Show that for a point set $P \in \mathbb{R}^2$, there exists a point $p \in P$ such that p is a $\frac{1}{8}$-quadrant centerpoint of P.

   [8+4+8=20]

4.(a) Define VC-dimension of a set system (X, R).

   (b) Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of $d+2$ points in $\mathbb{R}^d$. Show that there exists two disjoint point sets C and D of P such that $CH(C) \cap CH(D) \neq \emptyset$ and $C \cup D = P$. Here $CH(P)$ denotes the convex hull of the point set P.

   (c) Let $P \subseteq \mathbb{R}^d$ be a finite point set; r be any point in the convex hull of P, and h^+ be a halfspace of $\mathbb{R}^d$ passing through r. Prove that there exists a point of P inside h^+.

   (d) Now prove that if $S = (X, \mathcal{R})$ be a set system with $X = \mathbb{R}^d$ and $\mathcal{R}$ is the set of all closed halfplanes in $\mathbb{R}^d$, then VC-dimension of S is $d+1$.

   [4+4+8+8=24]

P.T.O.

5.(a) Define ϵ-net.

(b) Given an ϵ and a set of ordered points $A = \{a_1, a_2, \ldots, a_n\}$ on the real line $\mathbb{R}$, choose an ϵ-net for ranges which are intervals on that real line.

(c) Given an ϵ and a set of points $A = \{a_1, a_2, \ldots, a_n\} \in \mathbb{R}^2$ in convex position, choose an ϵ-net for ranges which are convex polygons in $\mathbb{R}^2$.

(d) Let n and d ($d \geq 2$) be two positive integers, and ϵ is a real number satisfying $0 < \epsilon < 1$. Let $(X, \mathcal{R})$ be a range space of VC-dimension d. If $A \subseteq X$ with $|A| = n$, then prove that there exists an ϵ-net $\mathcal{N}$ of A with respect to $\mathcal{R}$ with $|\mathcal{N}| \leq \lceil \frac{d \log n}{\epsilon} \rceil$.

[4+4+4+12=24]

6. Given a point set P in $\mathbb{R}^2$ with even number of points, a halving segment $[p_i, p_j]$ is defined by two points $p_i, p_j \in P$ provided the line containing p_i, p_j partitions the point set $P \setminus \{p_i, p_j\}$ into two equal halves.

   (a) Show an instance of the point set with 12 points with maximum number of halving segments.

   (b) Prove that every point in P participates in at least one halving segments.

   (c) Characterize those points in P which participate in exactly one halving segment.

   (d) Show that the maximum number of halving segments in a point set is $O(|P|^{3/2})$.

[6+4+4+10=24]

INDIAN STATISTICAL INSTITUTE

End-Semestral Examination: 2019

Subject Name : **Advanced Cryptology** Date: 22nd April 2019
Course Name : M.Tech. (CS) II yr. Max Score: 50 Duration: 180 Mins
Note: Attempt all questions. Marks are given in brackets. Total score is 55. But maximum you can score is 50. Use separate page for each question.

1.[10] Let $f : \{0,1\}^{2n} \to \{0,1\}^n$ be a compression function. Find $O(1)$ memory and $O(2^{n/2})$ collision finding algorithm for f. Briefly justify the correctness of your algorithm.

2.[5] Let e_K be an n-bit block cipher and $f_L : \{0,1\}^n \to \{0,1\}^{(l+1)n}$ be a PRF. Given a message $M = (m_1, m_2, \ldots, m_l)$, $|m_i| = n$ and an n-bit nonce N, define an AE as

$$\mathcal{E}_{K_1,K_2,L}(N, M) = f_L(N) \oplus (e_{K_1}(m_1) \| \cdots \| e_{K_1}(m_l) \| e_{K_2}(m_1 \oplus \cdots \oplus m_l)).$$

Show that this AE construction is not secure. Write down your complexity of your attack algorithm.

3.[4 + 6 = 10] Characterize the function graph of an n-bit involution functions. Let π_1, π_2 be two random independent n-bit involutions. Show that the expected size of a cycle of $\pi_2 \circ \pi_1$ containing the point 0^n is $O(2^{n/2})$.
(You can use the fact that the expected number of fixed points of a random involution π is $O(2^{n/2})$.)

4.[3 + 4 + 8 = 15] Define (t, ϵ)-hardcore function. Given a (t, ϵ) one-way n-bit permutation f, let us define $f'(x, r) = (f(x), r)$, $|r| = n$. Show that f' is also (t', ϵ) one-way with $t' \approx t$. Show that the dot-product $\langle x, r \rangle$ is a hard-core function for f'.

5. [3 + 5 = 8] Define (t, ϵ)-m-multi-PRF. Show that if f is (t, ϵ)-PRF then it is $(t', m\epsilon)$-m-multi-PRF where $t' \approx t$.

6 [2+5=7] Define a (t, n) secret sharing scheme. Define Shamir's (t, n) secret sharing and show its correctness.

INDIAN STATISTICAL INSTITUTE

End-Semester Examination 2018-2019 | M. Tech. (CS) Second Year
Subject: VLSI Testing and Fault tolerance | Date: 23·04·2019
Maximum marks = 100 | Time: 3 hours
Answer as much as you can, the maximum you can score is 100

---

1. [8 + 10 + 7 = 25]

(a) Draw a Büchi automaton that accepts the following ω-regular language:

$L = \{\sigma \in \{A, B\}^{\omega} \mid \sigma$ contains ABA infinitely often, but AA only finitely often$\}$

(b) Briefly explain the fix-point computation algorithm for verifying a CTL property of the form **EG** φ on a given model.

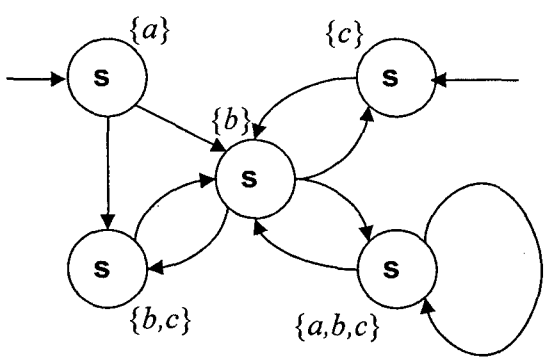
(c) Consider the teo properties (i) **A(p U q)** and (ii) $\neg$**E[**$\neg$**q U (**$\neg$**p** $\wedge$ $\neg$**q)]** $\wedge$ $\neg$**EG** $\neg$**q**

Are these equivalent? If so, provide a formal proof. If not, provide a counterexample.

2. [7 + (3 X 6) = 25]

(a) Let AP = {b}. Draw a *single* Kripke structure that satisfies the CTL property **AG AF (b)** and the LTL property **F**$\neg$**b.**

(b) Consider the following transition system, TS over AP ={ a, b, c }



Explain with brief justification for each LTL formula φ_i below, whether TS $\models \varphi_i$ holds. If TS does not satisfy φ_i then provide a path $\pi \in$ Paths(TS) such that π does not satisfy φ_i

$\varphi_1 = FG\ c$ $\quad$ $\varphi_2 = GF\ c$ $\quad$ $\varphi_3 = X\neg c \Rightarrow XX\ c$

$\varphi_4 = G\ a$ $\quad$ $\varphi_5 = a\ U\ G(b \vee c)$ $\quad$ $\varphi_6 = (XX\ b)\ U\ (b \vee c)$

3. [7 + 6 + (6 + 6) = 25]

(a) Compute the expression coverage for the following expression in a run that has encountered these values for a, b, and c: (1,1,0), (0,0,1), (1,0,1), and (0,1,0):

((b ? a : c) (a + b'c'))

(b) Briefly explain with an example, Hoskote's algorithm for computing the coverage of a specification with respect to a given design.

PTO

(c) Consider the following set of requirements for an intelligent traffic light controller that takes r1 and r2 as input, and produces g1 and g2 as output.

- Whenever r1 goes high, g1 must be asserted for the next two cycles
- When both r1 and r2 are low, the control assigns the grant on g2 in the next cycle
- The grant lines g1 and g2 are mutually exclusive

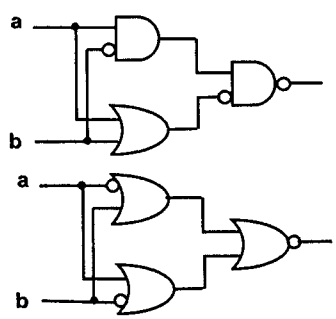
(i) Are the above properties consistent? Justify your answer.

(ii) Are the above properties complete? Justify your answer.

4. [(4 + 9 + 5) + 7 = 25]

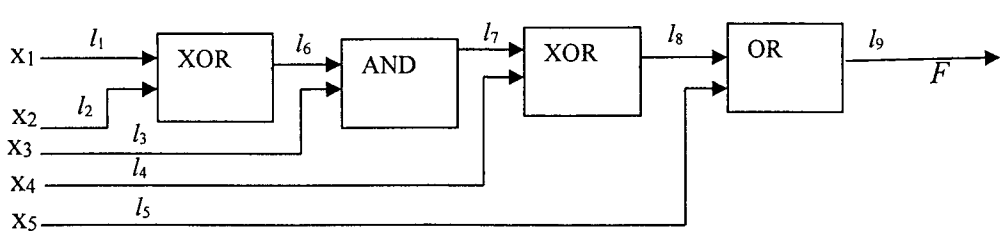
(a) State the register correspondence problem for equivalence checking between sequential circuits. Present an algorithm for solving the register correspondence problem and illustrate its working with an example. If no register correspondence exists, can we conclude that the sequential circuits are not equivalent? If not, give a counter-example.

(b) Perform SAT based equivalence checking to check if the two circuits shown below are equivalent.



5. [(6 + 6 + 6 + 7) = 25]

Consider the following circuit-under-test (CUT) in the figure below with 5 Boolean inputs x1, x2, x3, x4, x5, which consists of two XOR gates, one 2-input AND, one 2-input OR gate, producing a Boolean function F. The lines in the circuit are labeled as I1, I2, I3, ..., I9.



(a) Express the output Boolean function F in terms of the primary input variables.

(b) Determine a test vector that detects the fault I2 stuck-at 0, assuming no other faults exist.

(c) Ignore faults in the interior of XOR blocks and consider single stuck-at-0 and stuck-at-1 faults only on lines I1, I2, I3, ..., I9. Derive a complete test set T that gives 100% fault coverage. Justify your argument.

(d) Compute the Boolean difference of the function F with respect to line I4, and determine all the test vectors that are capable of detecting I4 stuck-at-1.

INDIAN STATISTICAL INSTITUTE

Second Semester Examination : 2018-2019

Course Name : M.TECH. (CS) YEAR II

Subject name : FORMAL METHODS FOR COMPUTER SCIENCE: SELECTED TOPICS

Date : 23.04.2019 Maximum Marks : 50 Duration : 3 hours

Answer any 8 questions. Notations are used as in the class. 2 marks are reserved for neatness.

1. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two Kripke models, such that $(\mathcal{M}_1, w_1)$ is bisimilar to $(\mathcal{M}_2, w_2)$. Show that for all basic modal formulas φ, $\mathcal{M}_1, w_1 \models \varphi$ iff $\mathcal{M}_2, w_2 \models \varphi$.

   Hence, or, otherwise, show that the unary operator $U\varphi$ defined by:

   $\mathcal{M}, w \models U\varphi$ iff for all $v \in W$, $\mathcal{M}, v \models \varphi$,

   is not definable in the basic modal language. [6]

2. Let $\mathcal{M}$ be a Kripke model and w be a world in it. Let φ be a basic modal formula. Consider the modal evaluation game $\mathcal{G}(\mathcal{M}, w, \varphi)$. Show that the player E has a winning strategy in $\mathcal{G}(\mathcal{M}, w, \varphi)$, if $\mathcal{M}, w \models \varphi$. [6]

3. Let φ be a consistent formula in LTL, and let $Cl(\varphi)$ denote the Fischer-Ladner closure of φ in LTL. Let $W = \{\Gamma \cap Cl(\varphi) : \Gamma \text{ is a maximal consistent set in } LTL\}$. Show the following:

   (a) If $x \subset Cl(\varphi)$, then $\vdash_{LTL} \hat{x} \rightarrow \bigvee_{w \in W, w \supseteq x} \hat{w}$, where $\hat{y}$ denotes the conjunction of all formulas in y.

   (b) If $w \in W$, then $\vdash_{LTL} \hat{w} \rightarrow X\big(\bigvee_{X^{-1}w \subseteq v} \hat{v}\big)$, where $X^{-1}w = \{\varphi \mid X\varphi \in w\}$. [6]

4. Let φ be a CTL formula and let $\mathcal{M}$ be a model for φ. Can we always use the usual filtration method to show that φ is satisfiable in a finite model? Justify your answer. [6]

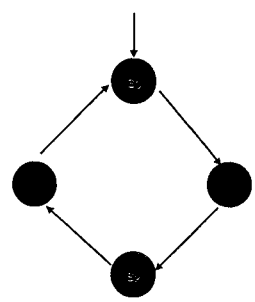
5. Second order logic with standard semantics is compact: Prove or disprove. [6]

6. Let $\Sigma = \{a, b\}$. Describe the languages expressed by the following $MSO(\Sigma)$ formulas and give a corresponding automaton:

   (a) $P_a(x) \vee P_b(x)$

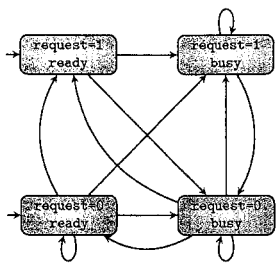
   (b) $\exists x(x \in X)$ [6]

7. Draw an ROBDD representation of the following transition system with proper justifications. [6]



PTO

8. Given a transition system TS and an LTL formula of the form $G\varphi$ with boolean φ, give an algorithm to check whether the TS satisfies $G\varphi$. Use the algorithm to check whether $G(p_1 \wedge \neg p_2)$ is satisfiable in the following TS, where the atomic propositions p_1 and p_2 denote 'request = 1' and 'status = busy', respectively. Here, 'request' is a boolean variable and 'status' takes up the value 'ready' or 'busy'. [6]

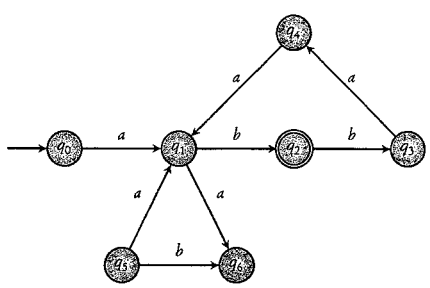


9. Consider the set of atomic propositions $AP = \{p, q\}$. Formulate the following properties as linear time properties over AP, and provide transition systems satisfying those properties.

   (a) every occurrence of p should eventually be followed by an occurrence of q.

   (b) p and q alternate infinitely often. [6]

10. Give an algorithm to check whether the language of a non-deterministic Büchi automaton is empty. Use the algorithm to show that the language of the following Büchi automaton is non-empty. [6]



11. Construct a Büchi automaton, providing proper justifications, for the LTL formula pUq, where p and q are atomic propositions. [6]

12. Given a deterministic Muller automaton will there always exist an equivalent deterministic Büchi automaton? Justify your answer. [6]

INDIAN STATISTICAL INSTITUTE

M. Tech. Computer Science Year II, 2018-19

Semester Examination

Advanced Pattern Recognition

DATE : 25.04.19

Maximum Marks: 100 Duration: 3 hours

Answers should be brief and to the point

A. Answer **any five** questions [5x3=15]

1. What can you infer about the performance of a binary classifier from ROC curve? 3
2. What is the main difference between probability and fuzzy membership value? 3
3. Explain roulette wheel selection strategy in genetic algorithms. 3
4. Which of the following points from each of the following sets of points below is an outlier? Explain why? 3
   i. (1-dimensional stationary data)
   { 1, 3, 2, 1, 3, 2, 75, 1, 3, 2, 2, 1, 2, 3, 2, 1 }
   ii. (2-dimensional stationary data)
   { (1, 9), (2, 9), (3, 9), (10, 10), (10, 3), (9, 1), (10, 2) }
   iii. (1-dimensional time series data)
   {1, 2, 3, 4, 3, 2, 1, 3, 73, 72, 74, 73, 74, 1, 2, 3, 4, 2}
5. State the properties of a Gram matrix. 3
6. How do you ensure diversity in an ensemble of classifiers? 3

B. Write short notes on **any five** [5x5=25]

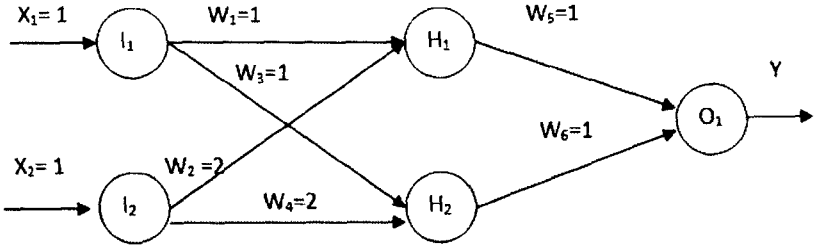
1. Lasso regression 5
2. Multi-view feature selection 5
3. Active learning 5
4. Fuzzy kNN 5
5. Kernel trick 5
6. Multi-label kNN 5

C. Answer **any three** questions [3x20=60]

1. Suppose you have to use a pre-trained deep neural network in a transfer learning scenario. How will you handle the following cases:
   i. The target dataset is small and similar to the source training dataset. 5
   ii. The target dataset is large and similar to the source training dataset. 5
   iii. The target dataset is small and different from the source training dataset. 5
   iv. The target dataset is large and different from the base training dataset. 5

2. Write the steps of BIRCH algorithm. 8
   Separate X = [[0, 1], [0.3, 1], [-0.3, 1], [0, -1], [0.3, -1], [-0.3, -1]] using BIRCH. 12
   [Assume branching factor 2 and threshold 0.5.]

3. Consider the following neural network with learning rate (eta) = 0.24, desired output = 1 and sigmoid activation function ($1/1+e^{-x}$).
   i. Perform one forward pass and calculate the error. 8
   ii. Calculate the updated weights for w_1, w_2, w_3, w_4, w_5 and w_6 using back-propagation. 12



4. In a data stream, transactions arrive continuously and the number of transactions can be potentially infinite.

   i. In this context consider a stream of data:
   9 3 5 2 7 1 6 5 8 4 9 1 5 8 4
   Use the damped window model with a damping factor of 0.1 to find the median and mean of the data stream. 15

   ii. Explain the meaning of "concept drift". 5

---

5. Assume that a sufficiently large language corpus is divided into six parts (r_1, r_2, ..., r_6). You observe the numbers of pronouns in three particular parts (say, in r_i, r_j and r_k) are 200, 100 and 300. The observed mean($\overline{n}$) is 200. For distribution of pronouns, let p_i denote the probability that a pronoun belongs to r_i and n_i denote the number of pronouns observed in r_i. Find one possible distribution (write down the values of p_i's and n_i's) of the pronouns in the six parts which is consistent with the following two constraints and the distribution should have **maximum entropy** over the other possible distributions:

$$\text{Constraint 1: } \overline{n} = \sum_{i=1}^{6} n_i p_i \text{ ; Constraint 2: } \sum_{i=1}^{6} p_i = 1$$

Justify your answer. **[10]**